Annals of Mathematics
Vol. 45, No. 3, July, 1944

ON FAMILIES OF REAL FUNCTIONS WITH A DENUMERABLE BASE

By Fritz Rothberger

(Received June 1, 1943)

Contents

§1. Statement of general problem.[1]

Definitions. A family $\mathfrak{F}$ of functions is said to have a *denumerable base* if there exists a sequence of functions $\{f_n(x)\}$ (not necessarily $\epsilon\, \mathfrak{F}$) such that any function $f \epsilon\, \mathfrak{F}$ is the limit of a subsequence of $\{f_n(x)\}$. The *domain* $\mathfrak{X}$ of a function $f(x)$ is the set of x's for which $f(x)$ is defined; we say $f(x)$ is a *function on* $\mathfrak{X}$. A *dyadic function* is a function taking only the values 0 and 1.

Notations. Throughout this paper, Italic capitals will denote sets of natural numbers (i.e., of positive integers), e.g., $A, B, X_i, \cdots$. In cases of doubt we shall write *"set of n.n."* for *"set of natural numbers."* Small Greek letters, except μ, ν, will denote transfinite ordinals, as usual.

Two sets of n.n., say A, B, will be said to be *equivalent*, $A \sim B$, if they differ in a finite number of elements only. Otherwise we have $A \nsim B$. Λ denotes the empty set. Thus "$E \sim \Lambda$" means "E is a finite set of n.n.," and "$E \nsim \Lambda$" means "$\overline{\overline{E}} = \aleph_0$." $A < B$ means A is *almost-contained* in B, i.e., $A - B \sim \Lambda$ (or: $A \subset B + \textit{finite set}$). $A > B$ means $B < A$. Obviously $A < B \cdot \& \cdot A > B : \equiv : A \sim B$. Two sets A and B are said to be *almost-disjoint*[2] if $A \cdot B \sim \Lambda$. The *complement* of A, i.e., the set of all n.n. *not* ϵA, is denoted by CA as usual.

The general problem might be stated as follows:

Problem: *Given two powers* $\mathfrak{m}$ *and* $\mathfrak{n}$, *does every family* $\mathfrak{F}$ *of power* $\mathfrak{m}$ *of functions on a set* $\mathfrak{X}$ *of power* $\mathfrak{n}$ *necessarily have a denumerable base?*

Obviously, we may suppose $\aleph_1 \leq \mathfrak{m}$, $\mathfrak{n} \leq 2^{\aleph_0}$, the other cases being trivial. For $\mathfrak{m} = \mathfrak{n} = \aleph_1$ the problem will be solved in the affirmative (§3). This settles the problem if the continuum hypothesis is assumed; *in this case there exists a denumerable base for the family of all functions of Baire class* 2, *in fact for all functions of the Baire-La Vallee-Poussin classification, etc.* (Cf. §3, corollary to Theorem 3). It is well known that such a base cannot consist of L-measurable functions, nor of functions having the Baire property.

[1] Cf., *Fund. Math.*, vol. 27 (1936) p. 293, *problème de M. Sierpiński*; also, W. Sierpinski, *Pont. Acad. Sci. Acta*, 4 (1940), p. 211. I am greatly indebted to Mr. A. S. Besicovitch for drawing my attention to this problem, in 1940; however, as I then left England very soon, we were not able to work on it in collaboration.

[2] Cf., A. Tarski, *Fund. Math.*, 12 (1928), p. 188–205.

Now we shall transform the problem. First of all, we can restrict ourselves to dyadic functions: this follows from

THEOREM 1. *Let $\mathfrak{F}$ be a family of real functions on $\mathfrak{X}$. If every family Φ of dyadic functions on $\mathfrak{X}$, such that $\overline{\overline{\Phi}} \leq \overline{\overline{\mathfrak{F}}}$, has a denumerable base, then $\mathfrak{F}$ has one.*

PROOF. We may suppose that $0 \leq f(x) \leq 1$ for every $f \in \mathfrak{F}$, throughout the domain $\mathfrak{X}$. Let

$$f(x) = \sum_{n=1}^{\infty} \frac{\varphi^{(n)}(x)}{2^n}, \qquad \varphi^{(n)}(x) = 0, 1. \tag{1.1}$$

Let Φ be the family of all dyadic functions $\varphi^{(n)}$ of (1.1) corresponding to any f of $\mathfrak{F}$. Then we have $\overline{\overline{\Phi}} \leq \aleph_0 \cdot \overline{\overline{\mathfrak{F}}}$, thus $\overline{\overline{\Phi}} \leq \overline{\overline{\mathfrak{F}}}$, and, by hypothesis, Φ has a base, say $g_1(x), g_2(x), \cdots g_n(x), \cdots$. We can assume these to be dyadic functions too. We shall show that the set of all functions of the form

$$\sum_{\nu=1}^{k} \frac{g_{n_\nu}(x)}{2^\nu} \tag{1.2}$$

where $(n_1, n_2, \cdots n_k)$ is any finite sequence of n.n., constitute a base for $\mathfrak{F}$.

Let $f \in \mathfrak{F}$, and let $\varphi^{(\nu)}$ be the corresponding dyadic functions (cf. (1.1)). For every ν there exists a sequence of n.n. $\{n_\mu^\nu\}(\mu = 1, 2, \cdots)$ such that $\varphi^{(\nu)}(x) = \lim_{\mu=\infty} g_{n_\mu^\nu}(x)$. Now, for $\mu > m$, we have

$$\sum_{\nu=1}^{\mu} \frac{g_{n_\mu^\nu}(x)}{2^\nu} = \sum_{\nu=1}^{m} \frac{g_{n_\mu^\nu}(x)}{2^\nu} + \epsilon, \text{ where } \epsilon < \frac{1}{2^m}, \text{ therefore } \lim_{\mu=\infty} \sum_{\nu=1}^{\mu} \frac{g_{n_\mu^\nu}(x)}{2^\nu} = \sum_{\nu=1}^{m} \frac{\varphi^{(\nu)}(x)}{2^\nu} + \epsilon,$$

and finally, $\lim_{\mu=\infty} \sum_{\nu=1}^{\mu} \frac{g_{n_\mu^\nu}(x)}{2^\nu} = f(x)$, so that $f(x)$ is the limit of functions of the form (1.2), q.e.d.

Again, our problem corresponds to a problem concerning families of sets of n.n.:

THEOREM 2. *Let $\mathfrak{F}$ be a family of power $\aleph_\xi$, consisting of the functions $f^\beta(x)$ on $\mathfrak{X}$, where $\beta = 1, 2, \cdots \omega, \cdots \mid \omega_\xi$; then $\mathfrak{F}$ has a denumerable base if and only if, for every $x \in \mathfrak{X}$ there exists a set (of n.n.) S_x, and for every $\beta < \omega_\xi$ a set $D^\beta \sim \Lambda$, such that*

$$\begin{cases} D^\beta < S_x & \text{if } f^\beta(x) = 0, \\ D^\beta < CS_x & \text{if } f^\beta(x) = 1. \end{cases} \quad (\text{for any } x \in \mathfrak{X}, \beta < \omega_\xi). \tag{1.21}$$

PROOF. If $\mathfrak{F}$ has a base of dyadic functions $f_n(x)$, there exists for any given β a sequence $\{n_\nu\}$ such that

$$f^\beta(x) = \lim_{\nu=\infty} f_{n_\nu}(x), \quad \text{for any } x \in \mathfrak{X}. \tag{1.3}$$

Let D^β be the set of these numbers n_ν, and put[3] $S_x = \underset{n}{E}(f_n(x) = 0)$. Now,

[3] By $\underset{n}{E}(\cdots)$ we denote the set of all n's satisfying the condition $(\cdots)$, cf., C. KURATOWSKI, *Topologie, I.*, Warszawa-Lwów, 1933.

it follows from (1.3) that

$$f_{n_\nu}(x) = \begin{cases} 0 & \textit{for almost all } n_\nu\textit{'s if } f^\beta(x) = 0, \\ 1 & \text{" \quad " \quad " \quad " \quad "} f^\beta(x) = 1. \end{cases}$$

Therefore, $D^\beta < S_x$ *if* $f^\beta(x) = 0$, *and* $D^\beta < CS_x$ *if* $f^\beta(x) = 1$.

On the other hand, if sets S_x and D^β satisfying (1.21) exist, put

$$f_n(x) = \begin{cases} 0 & \text{if } n \in S_x \\ 1 & \text{if } n \in CS_x. \end{cases}$$

Then $\{f_n(x)\}$ is a base, since $\lim_{n \in D^\beta} f_n(x) = \begin{cases} 0 \text{ if } f^\beta(x) = 0, \\ 1 \text{ if } f^\beta(x) = 1. \end{cases}$ q.e.d.

Corollary. *The relation (1.21) of theorem 2 implies that the D^β's are almost-disjoint sets, i.e. $D^\beta \cdot D^\gamma \sim \Lambda$ for any $\beta \neq \gamma$; provided that no two functions in the transfinite sequence $\{f^\beta(x)\}$ are identical.*

Proof. In this case we have $f^\beta(x) \neq f^\gamma(x)$ for some x if $\beta \neq \gamma$; suppose, for a particular value of x, that $f^\beta(x) = 0$, $f^\gamma(x) = 1$. Then, by (1.21), we have we have $D^\beta < S_x$, $D^\gamma < CS_x$, so that $D^\beta \cdot D^\gamma \sim \Lambda$, q.e.d.

As seen from the above theorems, the original problem is reduced to the question of the existence of certain families of sets of n.n. with—as we might say—"almost"-relations $\sim$, $<$, etc.

Before we can proceed any further, we have now to establish some lemmas about sets of sets (of n.n.) with such relations. (§2, below).

§2. Lemmas concerning sets of natural numbers

Lemma 1. *Given any finite or denumerable set of sets (of n.n.) $X_n \nsim \Lambda$, there exists a set A such that*

$$X_n \cdot A \nsim \Lambda \quad \textit{and} \quad X_n - A \nsim \Lambda (\textit{i.e., } X_n \cdot CA \nsim \Lambda), \quad \textit{for } n = 1, 2, \cdots$$

Proof. Let $Y_1, Y_2, \cdots Y_n, \cdots$ be a sequence (of sets) in which every X_n occurs infinitely many times, e.g., $Y_{2^n(2m-1)} = X_n$. Since we still have $Y_n \nsim \Lambda$ for $n = 1, 2, \cdots$, there exists a sequence of n.n., $a_1, b_1, a_2, b_2, \cdots a_n, b_n, \cdots$ no two terms of which are equal (i.e.: $a_m \neq a_n$, $a_n \neq b_n$, $a_m \neq b_n$, $b_m \neq b_n$), and such that

$$\begin{array}{c} a_1 \in Y_1, b_1 \in Y_1, \\ a_2 \in Y_2, b_2 \in Y_2, \\ \cdots\cdots\cdots\cdots \\ a_n \in Y_n, b_n \in Y_n, \\ \cdots\cdots\cdots\cdots \end{array} \tag{2.1}$$

Let A be the set of all the a's and B the set of all the b's, then (in view of the definition of the Y's), A and B have each infinitely many elements in common with every X_n: $A \cdot X_n \nsim \Lambda$, $B \cdot X_n \nsim \Lambda$. But besides we have $A \cdot B = \Lambda$, i.e. $B \subset CA$, so that $X_n \cdot CA \nsim \Lambda$, q.e.d.

DEFINITION. *A finite product (finite sum) is the product (sum) of a finite number of terms (not necessarily a finite set itself).*

LEMMA 2. *If $X_1, X_2, X_3, \cdots$ is a finite or infinite sequence of sets (of n.n.) such that every finite product $\prod_{\nu=1}^{m} X_{n_\nu} \nsim \Lambda$, then there exists a set $A \nsim \Lambda$ such that $A < X_n$, for any n.*

PROOF. If the sequence is finite let A be the product of all the X_n's. Otherwise, let A be a set of numbers $a_1, a_2, \cdots a_n, \cdots$ satisfying the following conditions: $a_1 \epsilon X_1$; $a_2 \epsilon X_1 \cdot X_2$, $a_1 \neq a_2$; $\cdots$; $a_n \epsilon \prod_{k=1}^{n} X_k$, $a_i \neq a_k$ for $i \neq k \leq n$; $\cdots$ (by hypothesis, the sets $X_1, X_1X_2, \cdots$ are all $\nsim \Lambda$). Thus almost all a_m's are contained in any particular X_n, so that $A < X_n$, q.e.d.

LEMMA 3. *Let $X_1, X_2, \cdots; Y_1, Y_2, \cdots$ be two (finite or infinite) sequences of n.n. such that, for any i, k, $X_i \cdot Y_k \sim \Lambda$. Then there exist two disjoint sets A, B (i.e. $A \cdot B = \Lambda$), such that $X_i < A$, $Y_k < B$ for any i, k.*

PROOF. If one of these sequences is finite, we may replace it by an infinite one, e.g., by filling it up with Λ's; thus we can assume both sequences to be infinite. Now we put

$$\begin{aligned} X'_1 &= X_1, & Y'_1 &= Y_1 - X_1, \\ X'_2 &= X_2 - Y_1, & Y'_2 &= Y_2 - (X_1 + X_2), \\ &\cdots\cdots & &\cdots\cdots \\ X'_n &= X_n - \textstyle\sum_{k=1}^{n-1} Y_k, & Y'_n &= Y_n - \textstyle\sum_{k=1}^{n} X_k, \\ &\cdots\cdots & &\cdots\cdots \end{aligned} \tag{2.2}$$

Then it is easily seen that

$$X'_i Y'_k = \Lambda \qquad \text{for any } i, k. \tag{2.3}$$

(Here we shall need the relation $=$, rather than $\sim$).

Since X_i and Y_k have only a finite number of elements in common (by hypothesis), we have, from (2.2)

$$X_n \sim X'_n, \qquad Y_n \sim Y'_n \qquad \text{for any } n. \tag{2.4}$$

Now, putting $A = \sum_{n=1}^{\infty} X'_n$, $B = \sum_{n=1}^{\infty} Y'_n$, it follows from (2.3) that $AB = \Lambda$, and from (2.4) that $X_n < A$, $Y_n < B$ (for any n), q.e.d.

Since $AB \sim \Lambda$ implies $B < CA$, we have

LEMMA 3^a. *Under the conditions of lemma 3 there exists a set A, such that $X_i < A$, $Y_k < CA$, for any i, k.*

Replacing the Y_k's by their complements, we find:

LEMMA 3^b. (Cf., Hausdorff's "Erster Einschaltungssatz"[4]). *Given two (finite or infinite) sequences of sets, $X_1, X_2, \cdots$ and $Y_1, Y_2, \cdots$, such that $X_i < Y_k$ for any i, k, there exists a set A such that $X_i < A < Y_k$ for any i, k.*

---

[4] F. HAUSDORFF, *Fund. Math.* 26 (1936), p. 243–4.

§3. Main theorem: Existence of a base for any family of power $\aleph_1$ of functions on a domain of power $\aleph_1$.

THEOREM 3. *Every family of power $\aleph_1$ of real functions defined on a domain of power $\aleph_1$ has a denumerable base.*

DEFINITION. *A function $f(x)$ whose domain is the set of all real numbers $-\infty < x < +\infty$ is called a function of a real variable.*

An immediate consequence of theorem 3 is the following

COROLLARY. *The continuum hypothesis implies the proposition: Every family of power of the continuum of real functions of a real variable has a denumerable base.*

It follows from theorems 1 and 2 that theorem 3 is equivalent to the existence of sets S_x and $D^\beta \nsim \Lambda$ satisfying (1.21) where $\overline{\overline{\mathfrak{X}}} = \aleph_1$, $\beta < \Omega$ and the $f^\beta(x)$'s are any given dyadic functions. Let $x_1, x_2, \cdots x\omega, \cdots x_\alpha, \cdots (\alpha < \Omega)$ be the elements of $\mathfrak{X}$, and put

$$S^0_\alpha = S_{x_\alpha}, \qquad S^1_\alpha = CS_{x_\alpha}, \qquad \text{for } \alpha < \Omega. \tag{3.1}$$

Then (1.21) may be written as follows:

$D^\beta < S^i_\alpha$, where $i = f^\beta(x_\alpha)$, for any $\alpha, \beta < \Omega$; and we may suppose (cf. Corollary of Theorem 2) that $D^\beta \cdot D^\gamma \sim \Lambda$ for $\beta \neq \gamma$.

Thus finally, Theorem 3 is equivalent to the following theorem which we are going to prove.

THEOREM 4. *$\mathcal{F}$ being any family of dyadic functions $f^\beta(x_\alpha)$ $(\alpha, \beta < \Omega)$, there exists a family of sets (of n.n.) S^i_α and D^β, where $i = 0, 1$, and $\alpha, \beta < \Omega$, satisfying the following conditions:*

$$\left.\begin{array}{ll} (4.0) & S^0_\alpha = CS^1_\alpha, \\ (4.1) & D^\beta < S^i_\alpha, \quad \text{where} \quad i = f^\beta(x_\alpha), \\ (4.2) & D^\beta \nsim \Lambda, \quad \text{and} \quad D^\beta \cdot D^\gamma \sim \Lambda \quad \text{for} \quad \beta \neq \gamma, \end{array}\right\} \text{for any } \alpha, \beta, \gamma < \Omega.$$

PROOF. Let $\mathcal{F}$ be a family of dyadic functions $f^\beta(x_\alpha)$ where $\alpha, \beta < \Omega$. Using a method of transfinite induction, we shall prove the existence of a family of sets satisfying (4.0), (4.1) and (4.2), and besides the following condition:

$$\prod_{\nu=1}^{n} S^{i_\nu}_{\gamma_\nu} - \sum_{\mu=1}^{m} D^{\beta_\mu} \nsim \Lambda \quad \textit{where} \quad i_\nu = 0, 1; \tag{4.3}$$
$$\gamma_1 < \gamma_2 < \cdots \gamma_n < \Omega \quad \textit{and} \quad \beta_\mu < \Omega, \qquad (m, n \textit{ finite}).$$

This formula may need an explanation. Consider the two transfinite sequences:

$$\begin{cases} S^{i_1}_1, S^{i_2}_2, \cdots S^{i_\omega}_\omega, \cdots S^{i_\alpha}_\alpha, \cdots, & (i_\alpha = 0, 1) \\ CD^1, CD^2, \cdots CD^\omega, \cdots CD^\beta, \cdots, & \end{cases} \tag{4.4}$$

where the indices i_α are arbitrarily fixed. Then (4.3) means that any finite product of terms of (4.4) is $\nsim \Lambda$. It follows by Lemma 2 that, given any denumerable set of terms of (4.4), there always exists a set $A \nsim \Lambda$ almost-contained in all of them. This fact—as will be seen later on—will be essential for the transfinite induction by which we are now going to proceed.

The sets S_1^0, S_1^1, D^1 can easily be defined such that $S_1^1 = CS_1^0$, $D^1 < S_1^i$, where $i = f^1(x_1)$, $D^1 \nsim \Lambda$, $S_1^i - D^1 \nsim \Lambda$ $(i = 0, 1)$. Now suppose the sets S_α^i, D^β already defined for all $\alpha < \xi$, $\beta < \xi$ such that

$$\left.\begin{array}{lll} (4.0;\xi) & S_\alpha^1 = CS_\alpha^0, & \\ (4.1;\xi) & D^\beta < S_\alpha^i \quad \text{where} \quad i = f^\beta(x_\alpha), & \\ (4.2;\xi) & D^\beta \nsim \Lambda \quad \text{and} \quad D^\beta \cdot D^\gamma \sim \Lambda \quad \text{for} \quad \beta \neq \gamma & \end{array}\right\} \text{for any } \alpha < \xi, \beta < \xi, \gamma < \xi$$

$$(4.3;\xi) \quad \prod_{\nu=1}^{n} S_{\gamma_\nu}^{i_\nu} - \sum_{\mu=1}^{m} D^{\beta_\mu} \nsim \Lambda, \begin{cases} \text{for any } \gamma_1 < \gamma_2 < \cdots \gamma_n < \xi, \beta_\mu < \xi, \\ \qquad\qquad\qquad\qquad i_\nu = 0, 1, \\ \text{and any finite } m, n. \end{cases}$$

Then we shall define S_ξ^0, S_ξ^1 and D^ξ, satisfying the relations $(4.0; \xi + 1)$, $(4.1; \xi + 1)$, $(4.2; \xi + 1)$ and $(4.3; \xi + 1)$.

(Note that, when ξ is a limit number the relation $(4.i; \xi)$ is equivalent (for any $i = 0, 1, 2, 3$) to the relation: "$(4.i; \eta)$ *for any* $\eta < \xi$"; it will therefore be sufficient to verify these relations for ordinals of the first kind only, i.e. the relations of the form $(4.i; \xi + 1)$.)

Throughout the following argument the ordinal ξ will be supposed to be fixed.

We begin with the definition of S_ξ^i, but for this purpose we shall need some preliminaries.

LEMMA. *For every finite product*

$$(5.1) \qquad P = \prod_{\nu=1}^{n} S_{\gamma_\nu}^{i_\nu}, \quad \textit{where} \quad \gamma_1 < \gamma_2 < \cdots < \gamma_n < \xi \quad \textit{and } i_\nu = 0, 1,$$

there exists a set $Q \nsim \Lambda$ *such that* $Q < P - D^\beta$ *for any* $\beta < \xi$.

PROOF OF LEMMA. We have $\prod_{\mu=1}^{m} (P - D^{\beta_\mu}) = P - \sum_{\mu=1}^{m} D^{\beta_\mu} \nsim \Lambda$ because of $(4.3; \xi)$. Thus every finite product of terms $P - D^\beta$ $(\beta < \xi)$ is $\nsim \Lambda$ and the β's form a denumerable set, therefore, by Lemma 2, there exists a set $Q \nsim \Lambda$ such that $Q < P - D^\beta$ for any $\beta < \xi$, q.e.d.

Now the set of all products of the form (5.1) is at most denumerable. Let $P_1, P_2, \cdots P_n, \cdots$ be a (finite or infinite) sequence consisting of these products. Then there exists a sequence of corresponding sets $Q_1, Q_2, \cdots Q_n, \cdots$, such that

$$(5.2) \qquad Q_n \nsim \Lambda, \qquad Q_n < P_n - D^\beta, \qquad \text{for any } \beta < \xi \text{ and any } n.$$

Therefore

$$(5.3) \qquad Q_n \cdot D^\beta \sim \Lambda \qquad \text{for any } \beta < \xi \text{ and any } n,$$

whence, by Lemma 3[a], there exists a set A such that

$$(5.4) \qquad Q_n < A \qquad \text{for any } n,$$

and

$$(5.5) \qquad D^\beta < CA \qquad \text{for any } \beta < \xi.$$

Besides, by Lemma 1 and (5.2), there exists a set E such that

(5.6) $$Q_n \cdot E \nsim \Lambda, \qquad Q_n \cdot CE \nsim \Lambda \qquad \text{(for any } n),$$

whence, by (5.4)

(5.7) $$Q_n \cdot AE \nsim \Lambda, \qquad Q_n \cdot A \cdot CE \nsim \Lambda \qquad \text{(for any } n).$$

Now, by (4.2; ξ) and Lemma 3^a, there exists a set H such that

(5.8) $$\left.\begin{aligned} D^\beta &< H \quad \text{if} \quad f^\beta(x_\xi) = 0, \\ D^\beta &< CH \quad \text{if} \quad f^\beta(x_\xi) = 1, \end{aligned}\right\} \text{(for any } \beta < \xi)$$

consequently, and from (5.5), we have

(5.9) $$\left.\begin{aligned} D^\beta &< H \cdot CA \quad \text{if} \quad f^\beta(x_\xi) = 0, \\ D^\beta &< CH \cdot CA \quad \text{if} \quad f^\beta(x_\xi) = 1. \end{aligned}\right\} \text{(for any } \beta < \xi)$$

Now we define the sets S_ξ^i as follows:

(6.1) $$\begin{cases} S_\xi^0 = H \cdot CA + AE, \\ S_\xi^1 = CH \cdot CA + A \cdot CE. \end{cases}$$

It is easily seen that

(6.2) $$S_\xi^1 = CS_\xi^0,$$

which proves (4.0; $\xi + 1$), and we shall show presently that S_ξ^i also satisfies the other conditions, at least as far as D^ξ is not involved.

By (5.9) and (6.1) we have

(6.3) $$D^\beta < S_\xi^i \qquad \text{if } i = f^\beta(x_\xi), \qquad \text{for any } \beta < \xi,$$

in accordance with (4.1; $\xi + 1$).

In order to show that

(6.4) $$S_\xi^i \prod_{\nu=1}^{k} S_{\gamma_\nu}^{i_\nu} - \sum_{\mu=1}^{m} D^{\beta_\mu} \nsim \Lambda$$
$$\text{(for any } \gamma_1 < \gamma_2 < \cdots \gamma_k < \xi, \beta_\mu < \xi; i, i_\nu = 0, 1 \text{ and any finite } k, m)$$

(cf., (4.3; $\xi + 1$)) it is sufficient to show that

(6.5) $$S_\xi^i \cdot Q_n \nsim \Lambda \qquad \textit{for } i = 0, 1, \textit{and any } n,$$

since $S_\xi^i \cdot \prod_{\nu=1}^k S_{\gamma_\nu}^{i_\nu} - \sum_{\mu=1}^m D^{\beta_\mu} = S_\xi^i \cdot P_n - \sum_{\mu=1}^m D^{\beta_\mu} > S_\xi^i \cdot Q_n$ (where n depends on the γ's), by (5.2). Now, by (6.1) we have $S_\xi^0 > AE$ and $S_\xi^1 > A \cdot CE$, thus, from (5.7), $S_\xi^0 \cdot Q_n > Q_n \cdot AE \nsim \Lambda$ and $S_\xi^1 \cdot Q_n > Q_n \cdot A \cdot CE \nsim \Lambda$, which implies (6.5) and therefore (6.4).

Now we proceed to the definition of D^ξ. Consider the denumerable (or finite) system of sets:

(7.1) $$\begin{cases} S_1^{i_1}, S_2^{i_2}, \cdots S_\omega^{i_\omega}, \cdots S_\xi^{i_\xi}; \text{where } i_\alpha = f^\xi(x_\alpha), \text{for } \alpha < \xi + 1; \\ CD^1, CD^2, \cdots CD^\omega, \cdots CD^\beta, \cdots (\beta < \xi). \end{cases}$$

(containing the ξ th term in the first row only). By (6.4) every finite product of terms of (7.1) is $\nsim \Lambda$. Therefore, by Lemma 2, there exists a set $U \nsim \Lambda$ which is almost-contained in each set of (7.1):

$$U < S^i_\alpha \quad \text{where } i = f^\xi(x_\alpha), \quad \text{for any } \alpha < \xi + 1, \tag{7.2}$$

and

$$U < CD^\beta, \text{ i.e. } U \cdot D^\beta \sim \Lambda, \quad \text{for any } \beta < \xi. \tag{7.3}$$

Let D^ξ be a proper subset of U:

$$D^\xi \nsim \Lambda, \tag{7.4}$$

$$D^\xi < U, \tag{7.4.1}$$

$$U - D^\xi \nsim \Lambda. \tag{7.4.2}$$

It follows from (7.2) and (7.4.1) that

$$D^\xi < S^i_\alpha \textit{ where } i = f^\xi(x_\alpha), \quad \textit{for any } \alpha < \xi + 1,$$

which, together with (6.3) and (4.1; ξ), proves (4.1; $\xi + 1$).

Similarly, it follows from (7.3) and (7.4.1) that $D^\xi \cdot D^\beta \sim \Lambda$, for any $\beta < \xi$; in view of (4.2; ξ) and (7.4), this proves (4.2; $\xi + 1$).

The relation (4.0; $\xi + 1$) follows from (6.2), as mentioned before. Thus only (4.3; $\xi + 1$) remains to be proved.

For this it is obviously sufficient to show that

$$S^i_\xi \cdot \prod_{\nu=1}^{k} S^{i_\nu}_{\gamma_\nu} - \left(D^\xi + \sum_{\mu=1}^{m} D^{\beta_\mu}\right) \nsim \Lambda \tag{8.1}$$

for any finite k, m and any $\gamma_1 < \gamma_2 < \cdots \gamma_k < \xi$; $\beta_\mu < \xi$; $i, i_\nu = 0, 1$,

or, which is the same thing,

$$(S^i_\xi \cdot P_n - D^\xi) - \sum_{\mu=1}^{m} D^{\beta_\mu} \nsim \Lambda, \quad \text{for } i = 0, 1, \text{ any } n \text{ and any } \beta_\mu < \xi, \tag{8.2}$$

where P_n has the same meaning as before.

From (7.2) and (4.0; $\xi + 1$) we have

(8.3) *for any given* $\alpha < \xi + 1$ *and* $i = 0, 1$: *either* $U < S^i_\alpha$ *or* $U \cdot S^i_\alpha \sim \Lambda$.

Therefore,

(8.4) *for any* n, *and* $i = 0, 1$: *either* $U < S^i_\xi \cdot P_n$ *or* $U \cdot S^i_\xi \cdot P_n \sim \Lambda$.

In the first case we have, from (7.3) and (7.4.2),

$$(S^i_\xi \cdot P_n - D^\xi) - \sum_{\mu=1}^{m} D^{\beta_\mu} > U - D^\xi \nsim \Lambda \quad (\text{for any } n \text{ and } \beta_\mu < \xi).$$

In the second case we have, by (7.4.1), $S^i_\xi \cdot P_n \cdot D^\xi \sim \Lambda$; thus, by (6.4)

$$(S^i_\xi \cdot P_n - D^\xi) - \sum_{\mu=1}^{m} D^{\beta_\mu} \sim S^i_\xi \cdot P_n - \sum_{\mu=1}^{m} D^{\beta_\mu} \nsim \Lambda.$$

Therefore we have in any case (8.2), which proves (4.3; $\xi + 1$).

This completes the proof. Summing up, we have defined by transfinite induction the sets S^i_ξ and D^ξ for any $\xi < \Omega$, and have verified simultaneously the relations (4.i; ξ) ($i = 0, 1, 2, 3$) for any ξ of the *first kind*, which is sufficient. It follows that the relations (4.0), (4.1), (4.2), (4.3) are satisfied throughout, q.e.d.

§4. On a problem of S. Ulam[5]

We begin with two definitions:

DEFINITION 1. *Φ being any given family of sets, we denote by Φ_σ the family of all sets which are denumerable sums of sets belonging to Φ, and by Φ_δ the family of all sets which are denumerable products of sets of Φ (e.g., if Φ is the family of all closed sets, Φ_σ is the class of $\mathbf{F}_\sigma$-sets, or if Φ is the class of all open sets, Φ_δ is the class of $\mathbf{G}_\delta$-sets).*

DEFINITION 2. *Φ being any family of sets, let $\mathbf{B}(\Phi)$ be the smallest family of sets which contains Φ and is closed with respect to the operations σ and δ (taking denumerable sums, resp. products); i.e., $\mathbf{B}(\Phi)$ is the smallest "Borel system" over Φ.*

S. Ulam posed a problem[5] which may now, in a generalized form, be stated thus:

Let $\mathcal{K}$ be a family of pointsets, of power of the continuum (e.g., let $\mathcal{K}$ be the class of all analytic sets); does there exist a denumerable family $\mathfrak{D}$ of sets, such that $\mathcal{K} \subset \mathbf{B}(\mathfrak{D})$? (A *pointset* is a subset of a Euclidean space, in particular a set of real numbers).

If we assume the continuum hypothesis we can solve this problem in the affirmative, and we can even prove a little more: viz., that

$$\mathcal{K} \subset \mathfrak{D}_{\sigma\delta} \cdot \mathfrak{D}_{\delta\sigma}.$$

This is an immediate consequence of the following theorem:

THEOREM 5. *Let $\mathfrak{X}$ be a set of power $\aleph_1$ and let $\mathcal{K}$ be a family of power $\aleph_1$ or subsets of $\mathfrak{X}$. Then there exists a denumerable family of sets, $\mathfrak{D}$, such that $\mathcal{K} \subset \mathfrak{D}_{\sigma\delta} \cdot \mathfrak{D}_{\delta\sigma}$.*

PROOF. We recall that the *characteristic function* of a set is the function which is = 1 on the set, and = 0 elsewhere. We shall consider the characteristic functions of the subsets of $\mathfrak{X}$, and take $\mathfrak{X}$ as the domain of these functions. Then we have a one-to-one correspondence between the subsets of $\mathfrak{X}$ and the dyadic functions on $\mathfrak{X}$, and, by a well-known theorem, this correspondence is invariant

[5] *Fund. Math.* 30 (1938), p. 365, problème 74.

with respect to taking limits[6]; i.e., the limit-set of a sequence of sets corresponds to the limit of their characteristic functions, and vice versa.

Let $\mathcal{K}$ be a family of sets in accordance with the hypothesis of our theorem, and let $\mathcal{F}$ be the corresponding family of characteristic functions. Then, by theorem 3, $\mathcal{F}$ has a denumerable base (consisting of dyadic functions), say $\{f_n(x)\}$; let $\mathfrak{D}$ be the (denumerable) family of the sets of which the functions $f_n(x)$ are the characteristic functions. It follows then, from the properties of characteristic functions, that every set belonging to $\mathcal{K}$ is the limit-set of some sequence of sets from $\mathfrak{D}$.

Now, the limit-set of a sequence of sets can be represented as a product of sums—and also as a sum of products—of sets out of that sequence. Therefore, the family of all limit-sets of $\mathfrak{D}$ is contained in $\mathfrak{D}_{\sigma\delta} \cdot \mathfrak{D}_{\delta\sigma}$, whence the theorem follows immediately, q.e.d.

Ottawa, Ontario
Canada

[6] Cf., F. Hausdorff, *Mengenlehre*, 2. Aufl. (Berlin, 1927), p. 21; also, C. Kuratowski, *Topologie, I.* (Monografie matematyczne 4, Warszawa-Lwów, 1933), p. 70.

ANNALS OF MATHEMATICS
Vol. 45, No. 3, July, 1944

SINGULAR HOMOLOGY THEORY

BY SAMUEL EILENBERG

(Received October 25, 1943)

CONTENTS

INTRODUCTION

The theory of abstract complexes divides into two analogous parts, namely the theory of star finite and the theory of closure finite complexes.[1] In a star finite complex, the homology groups are topologized and are based on infinite chains, while the cohomology groups are discrete and are based on finite cochains. In a closure finite complex it is the other way around. The theories are so closely analogous that they need not be treated separately; the results of one automatically apply to the other. In a finite complex the two theories merge.

The construction of a homology theory for topological spaces usually involves the theory of abstract complexes combined with some approximation device. Here we distinguish two essentially distinct methods.

The *first method* is based on the study of continuous mappings $T: X \to P$ of the given topological space X into polyhedra P. In the *second method* the approximation process is reversed: mappings $T: P \to X$ of polyhedra P into the given space X are studied.

The first method is best illustrated by the Alexandroff-Čech process in which the space X is mapped into the nerves of its coverings.[2] The resulting homology groups are topologized while the cohomology groups are discrete. This suggests that the first method is an extension of the star finite process in a complex. Close examination of the recent work of Alexander[3] and Kolmogoroff[4] confirm this conclusion. The theory of Vietoris and the recent "regular cycles" of Steenrod[5] all belong to this method.

The second method leads to the "singular" or the "continuous" homology

---

[1] See S. Lefschetz, *Algebraic Topology*, Amer. Math. Soc. Colloquium Publ., Vol. 27 (1942), Chapter III.

[2] P. Alexandroff, Ann. of Math., 30 (1928), p. 101–187; E. Čech, *Fund. Math.*, 19 (1932), p. 149–183.

[3] J. W. Alexander, Ann. of Math., 37 (1936), p. 698–708.

[4] A. Kolmogoroff, Comptes Rendues Paris, 202 (1936), p. 1144–1146.

[5] N. E. Steenrod, Ann. of Math., 41 (1940), p. 833–851.

theory. Discrete homology groups and topologized cohomology groups are obtained. This is the application of the closure finite theory to a space.

Both the star finite and the closure finite approach to topological spaces are fully justified by their applications. However, for each type of problem only one of the two proves to be really suitable. Since the two methods give the same results only in very isolated cases (for instance, compact metric absolute neighborhood retracts), using the wrong method will result in theorems with many inessential hypotheses. The reason for trying to use the star finite approach in places where the closure finite one is more natural is due to the fact that the star finite process has been highly perfected while the closure finite (or the singular) process is still in a very rudimentary stage. This is all the more puzzling since the first method is less than twenty years old while the second is as old as topology itself.

The best treatment of the singular homology theory so far has been given by Lefschetz.[6] He defines a singular simplex in a space X as a pair (s, T) where s is an oriented simplex and $T:s \to X$ is a continuous mapping. If $B:s \to s'$ is a barycentric mapping of s onto another oriented simplex of the same dimension as s, then

$$(*) \qquad (s, T) \equiv \pm (s', TB^{-1})$$

where the sign is $+$ or $-$ according as B preserves or reverses the orientation. Following a suitable definition of boundary and incidence numbers, Lefschetz arrives at what he calls the "total singular complex" $\mathfrak{S}(X)$ of the space X. In this closure finite complex homologies, cohomologies and products can be constructed.

The main difficulty with using the complex $\mathfrak{S}(X)$ is that it is not a bona fide abstract complex. Unfortunately relation (*) causes elements of order 2 to appear in the group of chains,[7] while in an abstract complex the group of chains ought to be free. There is the possibility of leaving out the elements of order 2 as degenerate, but this would make the use of the complex $\mathfrak{S}(X)$ cumbersome.

The main purpose of this paper is, by adjunction of a new idea to Lefschetz's method, to give a precise and systematic treatment of the singular homology theory. Instead of using oriented simplices as in Lefschetz's approach we shall use simplices with *ordered* vertices. In addition, the barycentric mapping $B:s \to s'$, used in defining equivalence, is required to preserve the order of the vertices. This modification leads to a much larger total singular complex $S(X)$ which, however, in contrast to $\mathfrak{S}(X)$, is a closure finite complex proper, so that the whole closure finite theory can be applied without restrictions. This program, including the product theory and the invariance theorem, is carried through in Chapters III–V.

In Chapter VI we show some connections with the homotopy groups. Two

[6] S. Lefschetz, Bull. Amer. Math. Soc., 39 (1933), p. 124–129; also *Algebraic Topology*, p. 311. See also P. Alexandroff and H. Hopf, *Topologie I*, Berlin, 1935, Ch. VIII, 5.

[7] See S. Lefschetz, Bull. Amer. Math. Soc., 39 (1933), p. 128.

theorems previously published by the author[8] without proofs are treated in detail.

The idea of using simplexes with a definite order of vertices turned out also to be helpful in clearing up some obscure points in the product theory in simplicial complexes. Given a simplicial polyhedron P, we choose a definite orientation for each simplex of P and arrive at an abstract complex $k(P)$. In order to introduce the products, one usually has to pick an ordering for all the vertices of P. At first the products depend upon the chosen order and only when we pass to the products of homology and cohomology classes does the order become immaterial. This results in a certain lack of naturality and uniqueness. In Chapter II we define a much larger abstract complex $K(P)$ in which each simplex appears with all the possible orderings of its vertices; we then show that the homology theories of $k(P)$ and $K(P)$ are essentially the same, while the product theory in $K(P)$ has very desirable uniqueness features (Ch. V).

Chapter I was meant as a review of the theory of abstract complexes. The technique of chain transformations and chain homotopies is developed. Some of the concepts, like that of *chain equivalence*, are new.

Chapter I

Abstract complexes[9]

1. Closure finite complexes

Let K be a collection of abstract elements σ^q called *cells*. With each cell there is associated an integer q called the dimension of σ^q. To any two cells σ^{q+1}, σ^q there corresponds an integer $[\sigma^{q+1}:\sigma^q]$ called the *incidence number*. K will be called a *closure finite abstract complex* provided the incidence numbers satisfy the following conditions:

(1.1) Given σ^{q+1}, $[\sigma^{q+1}:\sigma^q] \neq 0$ for only a finite number of q-cells σ^q;

$$\text{Given } \sigma^{q+1} \text{ and } \sigma^{q-1}, \quad \sum_{\sigma^q} [\sigma^{q+1}:\sigma^q][\sigma^q:\sigma^{q-1}] = 0. \tag{1.2}$$

The q-cells σ^q are taken as free generators of an abelian group $C^q(K)$; the elements c^q of $C^q(K)$ will be called (finite) integral q-dimensional chains of K. The *boundary operator* ∂ is a homomorphism

$$\partial: C^q(K) \to C^{q-1}(K) \tag{1.3}$$

defined for each generator σ^q as

$$\partial\sigma^q = \sum_{\sigma^{q-1}} [\sigma^q:\sigma^{q-1}]\sigma^{q-1} \tag{1.4}$$

Condition (1.1) insures that the summation is finite and therefore that $\partial\sigma^q$ is a well defined $(q-1)$-chain. We verify that condition (1.2) is equivalent with

$$\partial\partial = 0 \tag{1.5}$$

[8] S. Eilenberg, Proc. Nat. Acad. Sc., U. S. A., 26 (1940), p. 563–565.

[9] General reference: S. Lefschetz, *Algebraic Topology*, Ch. III and IV.

Very often in defining a complex we define the boundary operation (1.3) and prove (1.5) before introducing the incidence numbers. The incidence numbers can then be defined by (1.4) and conditions (1.1) and (1.2) can be proved.

2. Homology groups

Let G be a discrete abelian group. The q-dimensional chains of K over G are the formal finite sums

$$c^q = \sum g_i \sigma_i^q, \qquad g_i \in G$$

They form an abelian group $C^q(K, G)$.[10] The boundary operation is a homomorphism

$$\partial : C^q(K, G) \to C^{q-1}(K, G)$$

defined as

$$\partial c^q = \sum g_i \partial \sigma_i^q .$$

Again we have $\partial\partial = 0$. The chains c^q with $\partial c^q = 0$ are called *cycles*; they form a subgroup $Z^q(K, G)$ of $C^q(K, G)$. Chains c^q such that $c^q = \partial c^{q+1}$ for some c^{q+1} are called *boundaries*; they form a subgroup $B^q(K, G)$ of $C^q(K, G)$. It follows from $\partial\partial = 0$ that every boundary is a cycle, so

$$B^q(K, G) \subset Z^q(K, G).$$

The discrete factor group

$$H^q(K, G) = Z^q(K, G)/B^q(K, G)$$

is the q-dimensional *homology group* of K over G. If G is the additive group I of all integers we will write $C^q(K)$, $Z^q(K)$, $B^q(K)$ and $H^q(K)$, omitting the symbol for the group G.

The elements of $H^q(K, G)$ are called *homology classes*. Each cycle c^q determines uniquely a homology class. Two cycles c_1^q, c_2^q in the same homology class are called *homologous* (notation: $c_1^q \sim c_2^q$). Clearly $c_1^q \sim c_2^q$ if and only if $c_1^q - c_2^q$ is a boundary.

3. Cohomology groups

Let G be a topological group.[11] The q-dimensional *cochains* of K over G are functions f^q which with each q-cell σ^q in K associate an element $f^q(\sigma^q)$ of G. The q-cochains form an abelian topological group $C_q(K, G)$.[12] Since the q-cells σ^q are

---

[10] More precisely $C^q(K, G)$ is the group generated by symbols $g\sigma^q$ with relations $(g_1 + g_2)\sigma^q = g_1\sigma^q + g_2\sigma^q$.

[11] A topological group is a group with a topology with respect to which the group operations are continuous. No separation axioms are postulated, so that the factor group by a non-closed subgroup still may be regarded as a topological group.

[12] $C_q(K, G)$ is topologized as follows: Given a q-cell σ^q and an open set V in G a neighborhood (σ^q, V) in $C_q(K, G)$ is obtained by taking all the cochains f^q such that $f^q(\sigma^q) \in V$.

the free generators of the discrete group $C^q(K)$ each cochain leads to a unique homomorphism

$$f^q: C^q(K) \to G$$

and we have

$$C_q(K, G) = \mathrm{Hom}\{C^q(K), G\}.\text{[13]}$$

This second point of view will prevail in the sequel. The coboundary operator δ is a homomorphism

$$\delta: C_q(K, G) \to C_{q+1}(K, G)$$

defined as

$$(\delta f^q)c^{q+1} = f^q(\partial c^{q+1})$$

for $f^q \in C_q(K, G)$, $c^{q+1} \in C^{q+1}(K)$. The cochains f^q with $\delta f^q = 0$ are called *cocycles* and form a subgroup $Z_q(K, G)$ of $C_q(K, G)$. Cochains f^q of the form $f^q = \delta g^{q-1}$ are called *coboundaries* and form a subgroup $B_q(K, G)$ of $C_q(K, G)$. Since

$$(\delta\delta f^q)c^{q+2} = (\delta f^q)(\partial c^{q+2}) = f^q(\partial\partial c^{q+2}) = f^q(0) = 0$$

we have $\delta\delta = 0$ and B_q is a subgroup of Z_q. The factor group

$$H_q(K, G) = Z_q(K, G)/B_q(K, G)$$

is the q^{th} *cohomology group* of K over G. It carries a topology[14] induced by that of $C_q(K, G)$. The elements of $H_q(K, G)$ are called *cohomology classes*. Each cocycle f^q determines uniquely a cohomology class. Two cocycles f_1^q, f_2^q in the same cohomology class are called *cohomologous* (notation $f_1^q \backsim f_2^q$). Clearly $f_1^q \backsim f_2^q$ if and only if $f_1^q - f_2^q$ is a coboundary.

4. Chain transformations

Let K_1 and K_2 be two abstract closure finite complexes and let τ be a collection of homomorphisms, one for each dimension q

$$\tau: C^q(K_1) \to C^q(K_2). \tag{4.1}$$

We say that τ is a *chain transformation*

$$\tau: K_1 \to K_2$$

if

$$\tau\partial = \partial\tau. \tag{4.2}$$

[13] By Hom $\{H, G\}$ we denote the additive group of all homomorphisms $\varphi: H \to G$ with G abelian. A topology in Hom $\{H, G\}$ is obtained by taking a compact subset X of H and a neighborhood V of 0 in G and considering the set of all φ such that $\varphi(X) \subset V$ as a neighborhood of 0 in Hom $\{H, G\}$.

[14] Since B_q need not be a closed subgroup of Z_q, the group H_q may not satisfy any separation axioms. Cf. footnote 11.

More specifically, consider the diagram

$$\begin{array}{ccc} C^q(K_1) & \xrightarrow{\tau} & C^q(K_2) \\ \downarrow \partial & & \downarrow \partial \\ C^{q-1}(K_1) & \xrightarrow{\tau} & C^{q-1}(K_2) \end{array}$$

Condition (4.2) means that the two mappings of $C^q(K_1)$ into $C^{q-1}(K_2)$ that can be derived from the diagram are equal.

The homomorphisms (4.1) can be extended to homomorphisms

$$\tau : C^q(K_1, G) \to C^q(K_2, G)$$

by putting

$$\tau(\sum g_i \sigma_i^q) = \sum g_i \tau(\sigma_i^q).$$

Condition (4.2) being still valid, we obtain

$$\tau : Z^q(K_1, G) \to Z^q(K_2, G), \qquad \tau : B^q(K_1, G) \to B^q(K_2, G);$$

consequently τ induces homomorphisms

$$\tau : H^q(K_1, G) \to H^q(K_2, G)$$

of the homology groups of K_1 into those of K_2.

For the cochains we have induced homomorphisms

$$\tau^* : C_q(K_2, G) \to C_q(K_1, G)$$

defined by

$$(\tau^* f^q)(c^q) = f^q(\tau c^q)$$

for $f^q \in C_q(K_2, G)$, $c^q \in C^q(K_1)$. We verify that

$$\tau^* \delta = \delta \tau^*$$

and therefore

$$\tau^* : Z_q(K_2, G) \to Z_q(K_1, G) \qquad \tau^* : B_q(K_2, G) \to B_q(K_1, G)$$

and consequently τ induces homomorphisms

$$\tau^* : H_q(K_2, G) \to H_q(K_1, G)$$

of the cohomology groups of K_2 into those of K_1. These homomorphisms will be called the dual homomorphisms induced by τ.

Given two chain transformations

$$\tau_1 : K_1 \to K_2, \qquad \tau_2 : K_2 \to K_3$$

the composite chain transformations

$$\tau_2 \tau_1 : K_1 \to K_3$$

is defined by $\tau_2\tau_1(c^q) = \tau_2[\tau_1(c^q)]$. We notice that for the induced dual homomorphisms we have $(\tau_2\tau_1)^* = \tau_1^*\tau_2^*$.

Let K be a complex and K_1 a collection of cells of K. We shall say that K_1 is a closed subcomplex whenever $\sigma^q \in K$ and $[\sigma^q:\sigma^{q-1}] \neq 0$ implies $\sigma^{q-1} \in K$. This condition is equivalent with the condition that for every chain in K_1 the boundary also is in K_1. If K_1 is a closed subcomplex we define a chain transformation

$$\epsilon: K_1 \to K$$

by setting $\epsilon(c^q) = c^q$ for every chain c^q in K_1.

The dual homomorphisms of the cochains

$$\epsilon^*: C_q(K, G) \to C_q(K_1, G)$$

can be described as follows. Given a q-cochain f^q in K i.e. given a homomorphism

$$f^q: C^q(K) \to G$$

the homomorphism

$$\epsilon^* f^q: C^q(K_1) \to G$$

is obtained by considering f^q on the subgroup $C^q(K_1)$ of $C^q(K)$.

5. Chain homotopy

Let two chain transformations

$$\tau_1: K_1 \to K_2 \qquad \tau_2: K_1 \to K_2$$

be given. A collection D of homomorphisms

$$D: C^q(K_1) \to C^{q+1}(K_2)$$

(one for each q) will be called a chain homotopy between τ_1 and τ_2 provided

$$\partial D c^q = \tau_2 c^q - \tau_1 c^q - D\partial c^q \tag{5.1}$$

for all $c^q \in C^q(K_1)$.

If such a chain homotopy D exists we say that τ_1 and τ_2 are *chain homotopic* (notation: $\tau_1 \simeq \tau_2$). It is clear that the relation $\simeq$ is symmetric, reflexive and transitive.

It follows from (5.1) that if c^q is a cycle then $\tau_2 c^q$ and $\tau_1 c^q$ are homologous; consequently we have

(5.2) *Chain homotopic transformations induce identical homomorphisms of the homology groups.*

The homomorphisms D induce dual homomorphisms

$$D^*: C_q(K_2, G) \to C_{q-1}(K_1, G)$$

defined as

$$(D^* f^q)c^{q-1} = f^q(Dc^{q-1})$$

for $f^q \in C_q(K_2, G)$, $c^{q-1} \in C^{q-1}(K_1)$. We verify that

$$\delta D^* f^q = \overset{*}{\tau_2} f^q - \overset{*}{\tau_1} f^q - D^* \delta f^q.$$

Consequently if f^q is a cocycle then $\overset{*}{\tau_2} f^q$ and $\overset{*}{\tau_1} f^q$ are cohomologous. Hence we get

(5.3) *Chain homotopic transformations induce identical dual homorphisms of the cohomology groups.*

We further remark that if only one of the chain transformations, say τ_1, and the homotopy operator D are given, then the other chain transformation could be defined using (5.1) as

$$\tau_2 c^q = \tau_1 c^q + \partial D c^q + D \partial c^q.$$

τ_2 will automatically be a chain transformation since

$$\begin{aligned} \partial \tau_2 c^q &= \partial \tau_1 c^q + \partial\partial D c^q + \partial D \partial c^q \\ &= \tau_1 \partial c^q + [\tau_2 \partial c^q - \tau_1 \partial c^q - D \partial\partial c^q] \\ &= \tau_2 \partial c^q. \end{aligned}$$

Clearly τ_1 and τ_2 will be chain homotopic with D as the homotopy operator.

6. Chain equivalences

Let K_1 and K_2 be two closure finite abstract complexes. We will denote by the symbol 1 the identity chain transformation

$$1 c^q = c^q$$

of K_1 into itself or of K_2 into itself.

Two chain transformations

$$\tau : K_1 \to K_2 \qquad \rho : K_2 \to K_1$$

will be said to form an *equivalence pair* if and only if

$$\rho\tau \simeq 1 \qquad \text{and} \qquad \tau\rho \simeq 1.$$

Each of the transformations τ, ρ will be then called a *chain equivalence.*

The homomorphisms τ and ρ induced on the homology groups are then inverses of one another; hence they are isomorphisms. The same holds for the dual homomorphisms τ^* and ρ^* of the cohomology groups. It follows that

(6.1) *A chain equivalence induces isomorphisms of the homology groups and of the cohomology groups of the two complexes involved.*

(6.2) *Let $\tau : K_1 \to K_2$ be a chain equivalence. The chain transformation $\rho : K_2 \to K_1$ such that τ and ρ form an equivalence pair is determined uniquely up to chain homotopy.*

In fact, if $\tau_1\rho$ amd $\tau_1\rho'$ are both equivalence pairs then

$$\rho \simeq \rho'\tau\rho \simeq \rho'$$

because $\rho'\tau \simeq 1$ and $\tau\rho \simeq 1$.

(6.3) *If τ, ρ is an equivalence pair and*

$$\tau \simeq \tau' \quad \text{and} \quad \rho \simeq \rho'$$

then τ', ρ' is also an equivalence pair.

The proof is obvious. The previous two propositions show that the concept of an equivalence pair applies to chain homotopy classes of chain transformations, and that then the elements of the pair determine each other uniquely, and could justly be called inverses of one another.

7. The Kronecker index

Let G_1, G_2 and G be three topological (abelian) groups. We shall say that G_1 and G_2 are paired to G provided a multiplication g_1g_2 for $g_1 \in G_1$, $g_2 \in G_2$ with values in G is given which is bilinear and continuous in each variable separately.

If we denote by I the additive group of all integers then G and I are always paired to G in the obvious manner.

Let G_1 and G_2 be paired to G with G_2 discrete and let

$$f^q \in C_q(K, G_1), \qquad c^q \in C^q(K, G_2)$$

be given with $c^q = \sum g_i\sigma_i^q$. We define the Kronecker index by setting:

$$KI(f^q, c^q) = \sum f^q(\sigma_i^q)g_i.$$

We verify that KI establishes a pairing of $C_q(K, G_1)$ and $C^q(K, G_2)$ to G. We further verify that

$$KI(f^q, \partial c^{q+1}) = KI(\delta f^q, c^{q+1}). \tag{7.1}$$

This implies immediately that

$$KI(\text{cocycle, boundary}) = 0$$
$$KI(\text{coboundary, cycle}) = 0$$

and therefore KI establishes a pairing of the groups $H_q(K, G_1)$ and $H^q(K, G_2)$ to G.

An important property of the Kronecker index is connected with chain transformations $\tau: K_1 \to K_2$. Let

$$f^q \in C_q(K_2, G_1) \qquad c^q \in C^q(K_1, G_2)$$

then

$$KI(f^q, \tau c^q) = KI(\tau^* f^q, c^q). \tag{7.2}$$

A most important case of pairing is that of a discrete group G with its character group[15] Char G, given by

$$\chi g = \chi(g) \qquad\qquad \chi \in \text{Char } G,\ g \in G,$$

[15] Char G = Hom $\{G, R_1\}$ where R_1 is the additive group of real numbers reduced mod 1. It is known that Char G is compact (or discrete) if G is discrete (or compact). Cf. Lefschetz, *Algebraic Topology*, Ch. II.

the values of the pairing are in the additive group R_1 of reals reduced mod 1. It can be shown easily that to every character

$$\chi \in \operatorname{Char} C^q(K, G)$$

there is one and only one cochain $f^q \in C_q(K, \operatorname{Char} G)$ such that

$$\chi(c^q) = KI(f^q, c^q).$$

This way an isomorphism

$$\operatorname{Char} C^q(K, G) \approx C_q(K, \operatorname{Char} G) \tag{7.3}$$

is obtained. This isomorphism leads in a well known fashion to the isomorphism

$$\operatorname{Char} H^q(K, G) \approx H_q(K, \operatorname{Char} G) \tag{7.4}$$

which is the homology-cohomology duality theorem.

CHAPTER II

SIMPLICIAL COMPLEXES

8. The complexes $k(P)$ and $K(P)$

Let P be a locally finite polyhedron with a fixed simplicial decomposition.[16] After choosing a definite orientation for each simplex of P we define the complex $k(P)$ as follows. The cells σ^q of $k(P)$ are the oriented q-simplices of P (the 0-simplices are not oriented). The incidence numbers $[\sigma^q:\sigma^{q-1}]$ are defined in the usual way and are $+1$, -1 or 0 according to whether σ^{q-1} is a positively oriented face, a negatively oriented face or is not a face at all of σ^q. The resulting complex $k(P)$ is closure finite.

We shall now define another complex $K(P)$ as follows. A q-cell of $K(P)$ is an ordered array $v_0 \cdots v_q$ of vertices of P with possible repetitions and with the condition that all the vertices in question have to lie on one geometric simplex of P. A q-cell without repetition will be called *proper*, with repetitions, it will be called *degenerate*.

Next we define the boundary operation as[17]

$$\partial(v_0 \cdots v_q) = \sum_{i=0}^{q} (-1)^i v_0 \cdots \hat{v}_i \cdots v_q \tag{8.1}$$

for the cells. We extend this operation to a homomorphism

$$\partial : C^q(K(P)) \to C^{q-1}(K(P)).$$

We verify by a straightforward computation from (8.1) that

$$\partial\partial = 0.$$

[16] For more details see Lefschetz, *Algebraic Topology*, pp. 93–98.

[17] The dash over v_i indicates that v_i is omitted.

Using the operation ∂ we can define the incidence numbers and get an abstract closure finite complex $K(P)$.

Both complexes $k(P)$ and $K(P)$ will be taken "unaugmented." That means that there are no cells of dimension -1. Consequently every 0-chain is a cycle and for a connected P we have $H^0(k(P), G) \approx G$ and $H^0(K(P), G) \approx G$.

Given a cell σ^q in either of the two complexes $k(P)$ or $K(P)$ we denote by $|\sigma^q|$ the smallest geometric simplex of the decomposition of P, containing the cell σ^q.

9. The chain transformation α

We now define a chain transformation

$$\alpha: K(P) \to k(P) \tag{9.1}$$

as follows. Given a q-cell $v_0 \cdots v_q$ in $K(P)$ we distinguish two cases.

(1) If $v_0 \cdots v_q$ is improper we set

$$\alpha(v_0 \cdots v_q) = 0$$

(2) If $v_0 \cdots v_q$ is proper then it determines uniquely a q-cell σ^q in $k(P)$. The vertices $v_0 \cdots v_q$ determine a chain $[v_0 \cdots v_q]$ of $k(P)$

$$[v_0 \cdots v_q] = \pm \sigma^q$$

according as the orientation of σ^q agrees with the order $v_0 \cdots v_q$ or not. We define

$$\alpha(v_0 \cdots v_q) = [v_0 \cdots v_q].$$

This defines α for each cell in $K(P)$. We extend α by linearity to get homomorphisms

$$\alpha: C^q(K(P)) \to C^q(k(P)).$$

To show that these homomorphisms determine a chain transformation we have to show that

$$\partial\alpha = \alpha\partial. \tag{9.2}$$

It is sufficient to verify (9.2) for each cell of $K(P)$. If $v_0 \cdots v^q$ is a proper q-cell of $K(P)$ then

$$\begin{aligned}
\partial\alpha(v_0 \cdots v_q) &= \partial[v_0 \cdots v_q] = \sum(-1)^i[v_0 \cdots \hat{v}_i \cdots v_q] \\
&= \sum(-1)^i\alpha(v_0 \cdots v_i \cdots v_q) = \alpha(\sum(-1)^i v_0 \cdots \hat{v}_i \cdots v_q) \\
&= \alpha\partial(v_0 \cdots v_q)
\end{aligned}$$

If $v_0 \cdots v_q$ is improper with $v_k = v_l$ for $k < l$ then by definition $\partial\alpha(v_0 \cdots v_q) = 0$. We compute $\alpha\partial$

$$\alpha\partial(v_0 \cdots v_q) = \sum (-1)^i\alpha(v_0 \cdots \hat{v}_i \cdots v_q)$$

in this summation all the cells will be improper except perhaps for $i = k$ and $i = l$. Hence

$$\alpha\partial(v_0 \cdots v_q) = (-1)^k[v_0 \cdots \hat{v}_k \cdots v_q] + (-1)^l[v_0 \cdots \hat{v}_l \cdots v_q]$$

since $v_k = v_l$ we have

$$[v_0 \cdots \hat{v}_k \cdots v_q] = (-1)^{l-k+1}[v_0 \cdots \hat{v}_l \cdots v_q]$$

and therefore $\alpha\partial(v_0 \cdots v_q) = 0$ as desired.

THEOREM 9.1. *The chain transformation*

$$\alpha: K(P) \to k(P)$$

is a chain equivalence.

COROLLARY 9.2. *The chain transformation α induces isomorphisms of the homology groups of $K(P)$ onto those of $k(P)$.*

COROLLARY 9.3. *The chain transformation α induces dual isomorphisms α^* of the cohomology groups of $k(P)$ onto those of $K(P)$.*

Theorem 9.1 will be proved in §11.

10. The join

Let $\hat{P}$ be the polyhedron obtained from P by taking the join of P with a vertex v_0 outside of P. We may regard $\hat{P}$ as a cone with P as base and v_0 as vertex. Given any q-cell σ^q of $K(\hat{P})$ we denote by $v_0\sigma^q$ the $(q + 1)$-cell obtained by writing out the array of vertices defining σ^q and placing v_0 in front of them all. For each q-chain c^q of $K(\hat{P})$ the $(q + 1)$-chain v_0c^q of $K(\hat{P})$ is then well defined. A direct computation shows that

$$\partial(v_0c^q) = c^q - v_0(\partial c^q) \qquad \text{for } q > 0 \tag{10.1}$$

In particular if c^q is a cycle we have $c^q = \partial(v_0c^q)$ and hence $c^q \sim 0$. This proves

LEMMA 10.1. *The homology groups of the complex $K(\hat{P})$ vanish for dimensions $q > 0$.*

In particular a geometric simplex s may be regarded as a join of any of its faces with the opposite vertex. Consequently the symbol vc^q is well defined for any vertex v of s and any q-chain c^q of $K(s)$.

COROLLARY 10.2. *If s is a geometric simplex, the homology groups of $K(s)$ vanish for all dimensions $q > 0$.*

11. Proof of Theorem 9.1

In order to construct a chain transformation

$$\bar{\alpha}: k(P) \to K(P) \tag{11.1}$$

such that α, $\bar{\alpha}$ form an equivalence pair, we choose a definite ordering of the vertices of the polyhedron P. Each q-cell σ^q of $k(P)$ can then be written uniquely as

$$\sigma^q = \pm [v_0 \cdots v_q] \text{ with } v_0 < \cdots < v_q .$$

We define

$$\bar{\alpha}\sigma^q = \pm v_0 \cdots v_q.$$

It is obvious that $\bar{\alpha}\partial = \partial\bar{\alpha}$ and therefore $\bar{\alpha}$ is a chain transformation. Clearly

$$\alpha\bar{\alpha}\sigma^q = \sigma^q$$

which means that

(11.2) $$\alpha\bar{\alpha} = 1.$$

In order to prove that

(11.3) $$\bar{\alpha}\alpha \simeq 1$$

we will define a homotopy operator

$$D: C^q(K(P)) \to C^{q+1}(K(P))$$

subject to the following conditions

(11.4) $$\partial D c^q = c^q - \bar{\alpha}\alpha c^q - D\partial c^q$$

(11.5) $$D\sigma^q \subset |\sigma^q|.$$

We proceed by induction. Define $D\sigma^0 = 0$. Since $\bar{\alpha}\alpha\sigma^0 = \sigma^0$ conditions (11.4) and (11.5) are satisfied. Suppose that D has been defined for all dimensions less than q so that (11.4) and (11.5) hold. Consider

$$c^q = \sigma^q - \bar{\alpha}\alpha\sigma^q - D\partial\sigma^q.$$

Since $\bar{\alpha}\alpha\sigma^q \subset |\sigma^q|$ and $D\partial\sigma^q \subset |\sigma^q|$ we have

$$c^q \subset |\sigma^q|.$$

Further

$$\partial c^q = \partial\sigma^q - \partial\bar{\alpha}\alpha\sigma^q - \partial D\partial\sigma^q = \partial\sigma^q - \bar{\alpha}\alpha\partial\sigma^q - (\partial\sigma^q - \bar{\alpha}\alpha\partial\sigma^q - D\partial\partial\sigma^q) = 0.$$

Hence c^q is a cycle in $K(|\sigma^q|)$. Since by Corollary 10.2 every cycle in $K(|\sigma^q|)$ bounds in $K(|\sigma^q|)$ there is a $(q + 1)$-chain $D\sigma^q$ in $K(P)$ such that

$$D\sigma^q \subset |\sigma^q| \qquad \partial D\sigma^q = c^q.$$

This completes the definition of D and proves Theorem 9.1.

12. Proper and degenerate chains

We recall that a cell σ^q of $K(P)$ is called proper if its vertices do not contain repetitions, and degenerate otherwise. Accordingly a chain of $K(P)$ will be called *proper* or *degenerate* if it is composed of proper or degenerate cells only.

(12.1) *Each cycle in $K(P)$ is homologous to a proper cycle.*

(12.2) *Each degenerate cycle in $K(P)$ is bounding.*

Given a cycle c^q in $K(P)$ consider the cycle $\bar{\alpha}\alpha c^q$. Since $\bar{\alpha}\alpha \simeq 1$ we have $c^q \sim \bar{\alpha}\alpha c^q$. This proves (12.1) since $\bar{\alpha}\sigma^q$ is proper for any σ^q in $k(P)$. If c^q is degenerate then $\alpha c^q = 0$ and $\bar{\alpha}\alpha c^q = 0$. Hence $c^q \sim 0$.

The proper cells of $K(P)$ form a closed subcomplex $K_p(P)$. The inclusion relation induces homomorphic mappings of the homology groups of K_p into those of K. It follows from (12.1) that these homomorphisms are all mappings *onto*. We will show that generally they are *not* isomorphisms i.e. a cycle in K_p may bound in K without bounding in K_p. Let P be an interval with endpoints v_0, v_1. The chain

$$c^1 = v_0v_1 + v_1v_0$$

is then a proper cycle and

$$c^1 = \partial(v_0v_1v_0 - v_0v_0v_0).$$

However c^1 cannot be the boundary of proper 2-chain since $K(P)$ contains no proper 2-cells.

CHAPTER III

SINGULAR HOMOLOGIES IN A SPACE

13. The singular complex $S(X)$

Let s be a non-degenerate q-dimensional geometric simplex in some Euclidean space. If the vertices of s are given in a definite order $p_0 < p_1 < \cdots < p_q$ we shall say that s is an *ordered simplex* and write $s = < p_0 \cdots p_q >$. We shall denote by $s^{(i)}$ the face of s opposite the i^{th} vertex $s^{(i)} = < p_0, \cdots, \hat{p}_i, \cdots p_q >$. Taken with the same order of vertices as in s, the faces $s^{(i)}$ are ordered simplices.

Given two ordered q-simplices s_1 and s_2 we denote by B_{s_1,s_2} the barycentric mapping

$$B_{s_1,s_2}: s_1 \to s_2$$

preserving the order of the vertices. Clearly B_{s_1,s_2} is unique.

Let X be a topological space.[18] By a *singular* q-simplex in X we understand a continuous mapping

$$T: s \to X$$

of an ordered q-dimensional geometric simplex s.

Two singular q-simplices

$$T_1: s_1 \to X, \qquad T_2: s_2 \to X$$

are called *equivalent* (notation: $T_1 \equiv T_2$) provided

$$T_2B_{s_1,s_2} = T_1.$$

We verify that this relation is reflexive, symmetric, and transitive. Consequently the totality of all singular q-simplices in X is split into disjoint equivalence classes.

[18] The topological spaces considered here are of the most general type, with no separation axioms postulated.

We remark here that if $T: s \to X$ is a singular q-simplex, then given any ordered q-simplex s' there is a *unique* $T': s' \to X$ such that $T \equiv T'$.

Let $C^q(X)$ be the free abelian group generated by these equivalence classes. Alternatively, $C^q(X)$ may be defined as the group with the singular q-simplices in X as generators and $T_1 \equiv T_2$ as relations. The elements of the group $C^q(X)$ will be called the *integral singular q-chains in X*.

We now proceed with the definition of the boundary operator for singular chains. Given a singular q-simplex

$$T: s \to X, \qquad s = < p_0 \cdots p_q >,$$

consider the singular $(q - 1)$-simplices

$$T^{(i)}: s^{(i)} \to X$$

defined by[19] $T^{(i)} = T \mid s^{(i)}$. We define the boundary of T to be

$$\partial T = \sum_{i=0}^{q} (-1)^i T^{(i)}.$$

It is clear that if $T_1 \equiv T_2$ then $T_1^{(i)} \equiv T_2^{(i)}$ and therefore $\partial T_1 = \partial T_2$ in $C^{q-1}(X)$. Therefore we get a homomorphism

$$\partial: C^q(X) \to C^{q-1}(X).$$

We further verify that

$$\partial\partial = 0.$$

Consequently the boundary operation ∂ can be used to define incidence numbers and leads to a closure finite abstract complex that we will denote by $S(X)$ and call the *singular complex* of the space X. By definition we have

$$C^q(S(X)) = C^q(X).$$

The groups of the complex $S(X)$ will be called the singular groups of X:

$$H^q(X, G) = H^q(S(X), G) \qquad G \text{ discrete}$$

$$H_q(X, G) = H_q(S(X), G) \qquad G \text{ topological}$$

We notice that a cochain

$$f \in C_q(S(X), G)$$

can be considered either as a homomorphism

$$f: C^q(X) \to G$$

or as a function associating with each singular q-simplex T an element $f(T)$ of G, so that $f(T_1) = f(T_2)$ for $T_1 \equiv T_2$.

---

[19] Given a mapping $\varphi: X \to Y$ and a subset A of X we denote by $\varphi \mid A$ the mapping φ restricted to the subset A.

14. The invariance theorem

Let P be a simplicial polyhedron as described in §8. For P we have constructed three abstract closure finite complexes: $k(P)$, $K(P)$, and $S(P)$. We have already compared the complexes $k(P)$ and $K(P)$ using the chain equivalence $\alpha:K(P) \to k(P)$. We now proceed to compare the complexes $K(P)$ and $S(P)$.

We define a chain transformation

$$\beta:K(P) \to S(P) \tag{14.1}$$

as follows. Given a q-cell $v_0 \cdots v_q$ in $K(P)$, we consider an ordered q-dimensional simplex $s = < p_0 \cdots p_q >$ and a barycentric mapping $T:s \to P$ such that $T(p_i) = v_i$. Clearly T exists since $v_0, \cdots, v_q$ are all in the same simplex of P and T is unique. We set

$$\beta(v_0 \cdots v_q) = T$$

to get (14.1). It is clear that $\beta\partial = \partial\beta$ so that β is a chain transformation.

The reason the simplex s had to be constructed outside of P is because of the possibility of repetitions among the vertices $v_0, \cdots v_q$. If $v_0 \cdots v_q$ were non-degenerate we could have taken s to be contained in P and T to be the identity mapping.

One of the basic results of this paper is the following

THEOREM 14.1. *The chain transformation*

$$\beta:K(P) \to S(P)$$

is a chain equivalence.

The proof of this theorem requires preparatory considerations concerning barycentric subdivisions, and will be postponed until the next chapter.

This theorem jointly with Theorem 9.1 implies that the three complexes $k(P)$, $K(P)$ and $S(P)$ have isomorphic homology and cohomology groups. In particular since $S(P)$ is a topological invariant of P as a *space* it follows that the homology theories of $k(P)$ and $K(P)$ depend only upon P as a *space* and not upon the particular representation of P as a simplicial polyhedron.

15. Continuous mappings

Let

$$\varphi:X \to Y$$

be a continuous mapping of a space X into a space Y. Given a singular q-simplex in X

$$T:s \to X$$

we have the singular q-simplex in Y

$$\varphi T:s \to Y.$$

This leads to a chain transformation

(15.1) $$\varphi: S(X) \to S(Y).$$

If X is a simplicial polyhedron P and

$$\varphi: P \to Y$$

is a continuous mapping, then combining β with φ we obtain a chain transformation

$$\varphi\beta: K(P) \to S(Y).$$

This observation leads to the following convention. Let P be a simplicial polyhedron, c^q a q-chain in the complex $K(P)$ and $\varphi: P \to Y$ a continuous mapping. Then $\varphi\beta(c^q)$ is a q-chain in $S(Y)$, we shall write

$$(P, c^q, \varphi) = \varphi\beta(c^q).$$

This symbolism will prove to be advantageous in many proofs and applications.

16. Prisms

The singular homology theory is particularly useful in connection with problems dealing with homotopy. In order to prepare the ground for such applications we introduce a few definitions.

A *prism* of dimension $q + 1$ is the cartesian product $s \times I$ of an ordered q-simplex s and the closed interval $0 \leqq t \leqq 1$. The *bases* $s \times 0$ and $s \times 1$ of the prism $s \times I$ are q-simplices, in which the vertices are ordered just as in s. The *sides* $s^{(i)} \times I$ are prisms of dimension q.

We turn the prism into a simplicial polyhedron as follows. Let $s \times 0 = < p_0 \cdots p_q >$ and let $s \times 1 = < \bar{p}_0 \cdots \bar{p}_q >$ be the bases of $s \times I$. We subdivide $s \times I$ by considering the $(q + 1)$-simplexes

$$p_0 \cdots p_i \bar{p}_i \cdots \bar{p}_q$$

and all their faces. The simplicial subdivision of $s \times I$ so obtained will be called the *standard division*. Its main property is consistency; i.e., the standard divisions of the sides $s^{(i)} \times I$ agree with the division of $s \times I$. We also notice that the bases $s \times 0$ and $s \times 1$ are not subdivided.

In the complex $K(s \times I)$, with $s \times I$ taken in the standard division, we consider the $(q + 1)$-chain

$$d(s \times I) = \sum_{i=0}^{q} (-1)^i p_0 \cdots p_i \bar{p}_i \cdots \bar{p}_q$$

which we will call the *basic* $(q + 1)$*-chain on the prism* $s \times I$. We verify that

(16.1) $$\partial d(s \times I) = s \times 1 - s \times 0 - \sum_{i=0}^{q} (-1)^i d(s^{(i)} \times I)$$

A *singular* $(q + 1)$*-dimensional prism* in a space X is a continuous mapping

$$R: s \times I \to X$$

of a $(q + 1)$-prism $s \times I$ into X. The *bases*

$$R(0) = R \mid s \times 0 \qquad R(1) = R \mid s \times 1$$

are singular q-simplexes in X. The sides

$$R^{(i)} = R \mid s^{(i)} \times I$$

are singular q-prisms in X.

Each singular prism $R: s \times I \to X$ generates a singular $(q + 1)$-chain in X:

$$c(R) = (s \times I,\ d(s \times I),\ R)$$

where $s \times I$ is in its standard division and $d(s \times I)$ is the basic $(q + 1)$-chain of $K(s \times I)$. We verify at once that (16.1) implies

$$\partial c(R) = R(1) - R(0) - \sum_{i=0}^{q} (-1)^i c(R^{(i)}). \tag{16.2}$$

Given two ordered q-simplices s_1 and s_2 we have denoted by B_{s_1,s_2} the unique barycentric mapping $s_1 \to s_2$ preserving the order of the vertices. The mapping B_{s_1,s_2} obviously induces a mapping

$$\tilde{B}_{s_1,s_2}: s_1 \times I \to s_2 \times I.$$

Two singular prisms

$$R_1: s_1 \times I \to X, \qquad R_2: s_2 \times I \to X$$

will be called *equivalent* (notation $R_1 \equiv R_2$) provided

$$R_2 \tilde{B}_{s_1,s_2} = R_1 .$$

It is clear that if $R_1 \equiv R_2$ then $c(R_1) = c(R_2)$.

17. Homotopy

The constructions of the previous section will be used in the proof of the following

THEOREM 17.1 *If the continuous mappings $\varphi: X \to Y$ and $\psi: X \to Y$ are homotopic then the induced chain transformations $\varphi: S(X) \to S(Y)$ and $\psi: S(X) \to S(Y)$ are chain homotopic*

PROOF. Since φ and ψ are homotopic there is a mapping

$$\eta: X \times I \to Y$$

such that

$$\eta(x, 0) = \varphi(x) \qquad \eta(x, 1) = \psi(x).$$

Given a singular q-simplex of X

$$T: s \to X$$

we define a singular prism

$$R_T: s \times I \to Y$$

as follows

$$R_T(x, t) = \eta(T(x), t) \qquad x \in s, \qquad t \in I.$$

We notice that

$$R_T(0) = \varphi(T), \qquad R_T(1) = \psi(T)$$

hence putting $D(T) = c(R_T)$ we get by (16.2)

$$\partial D(T) = \psi(T) - \varphi(T) - D(\partial T).$$

This proves the theorem.

According to Hurewicz, two spaces X and Y are said to have the same *homotopy type* if there are two continuous mappings

$$\varphi: X \to Y \qquad \psi: Y \to X$$

such that $\psi\varphi$ is homotopic to the identity mapping of X into itself, and similarly $\varphi\psi$ is homotopic to the identity mapping of Y into itself. If we consider the induced chain transformations

$$\varphi: S(X) \to S(Y) \qquad \psi: S(Y) \to S(X)$$

we get from Theorem 17.1 that

$$\psi\varphi \simeq 1 \quad \text{and} \quad \varphi\psi \simeq 1$$

so that φ and ψ form a chain equivalence pair and the complexes $S(X)$ and $S(Y)$ are chain equivalent.

18. Relative homologies

We briefly outline the procedure leading to the homology and cohomology groups of a space X modulo a subset A. Every singular simplex in A may also be regarded as a singular simplex in X and this way the singular complex $S(A)$ may be considered a subcomplex of the larger complex $S(X)$. Clearly $S(A)$ is a closed subcomplex of $S(X)$ and therefore homology and cohomology groups of $S(X)$ mod $S(A)$ may be considered. These groups will be, by definition, the groups of the space X modulo the subset A.

Accordingly, a singular chain c^q in X will be called a cycle mod A if ∂c^q is a singular chain in A. We shall say that c^q bounds mod A if there is a singular chain c^{q+1} such that $c^q - \partial c^{q+1}$ is a chain in A. This way we get the homology groups of X mod A.

A cochain f^q in X will be called a cochain in X mod A if $f^q(T) = 0$ for every singular simplex T in A. Using cochains mod A throughout we arrive at the cohomology groups of X mod A.

With these definitions the properties of the Kronecker index and the duality between homology and cohomology are preserved.

If P is a simplicial polyhedron and Q a closed subpolyhedron then similar considerations can be applied to the complexes $k(Q)$ and $k(P)$ as well as the

complexes $K(Q)$ and $K(P)$. Since the chain transformation $\alpha: K(P) \to k(P)$ agrees on $K(Q)$ with the chain transformation $\alpha: K(Q) \to k(Q)$ and similarly for the chain transformation β, we conclude as before that the chain transformations α and β induce isomorphisms of the relative homology and cohomology groups.

CHAPTER IV

PROOF OF THE INVARIANCE THEOREM

19. V-sets

Let P' be the first barycentric subdivision of P. Given a point $p \in P$ let $St(p)$ be the union of all the simplices of P' containing p. The set $St(p)$ is closed, and we shall also consider its interior $\text{Int } St(p)$. Clearly $p \in \text{Int } St(p)$. For each point $p \in P$ we denote by $w(p)$ a vertex of P such that $p \in \text{Int } St(w(p))$. Obviously if p is a vertex of p then $p = w(p)$.

Given a simplex s of P we shall denote by $b(s)$ the barycenter of s and by $V(s)$ the union of the set s with the set $\text{Int } St[b(s)]$. For each $p \in V(s)$ the point $w(p)$ must be one of the vertices of the simplex s.

A subset A of P will be called a V-set provided $A \subset V(s)$ for some simplex s of P. Clearly every simplex of P is a V-set. Also every point of P has a neighborhood which is a V-set.

By standard methods, using the Lebesgue number we prove

LEMMA 19.1 *For each singular simplex*

$$T: s \to P$$

there is an integer n such that for each simplex s' of the n^{th} barycentric subdivision of s, the set $T(s')$ is a V-set.

20. The complex $S_V(P)$

We define a subcomplex $S_V(P)$ of $S(P)$ as follows. A singular simplex

$$T: s \to X$$

will be in $S_V(P)$ if and only if the set $T(s)$ is a V-set. Clearly $S_V(P)$ is a closed subcomplex of $S(P)$ and therefore the identity $\epsilon(T) = T$ is a chain transformation

$$\epsilon: S_V(P) \to S(P). \tag{20.1}$$

Since each simplex of P is V-set it follows that under the transformation

$$\beta: K(P) \to S(P) \tag{20.2}$$

the complex $K(P)$ is mapped into $S_V(P)$ and therefore we have a chain transformation

$$\beta_V: K(P) \to S_V(P) \tag{20.3}$$

such that

$$\beta = \epsilon\beta_V. \tag{20.4}$$

In order to prove that (20.2) is a chain equivalence it will therefore be sufficient to show that both (20.1) and (20.3) are chain equivalences.

We proceed to show that (20.3) is a chain equivalence. Given a singular q-simplex

$$T:s \to P \quad \text{in} \quad S_V(P)$$

where $s = \langle p_0 \cdots p_q \rangle$ is an ordered simplex, there is a simplex s_1 in P such that

$$T(s) \subset V(s_1).$$

For each point $p \in s$ the vertex $w[T(p)]$ is one of the vertices of s_1. This can be written as

$$w[T(s)] \subset s_1 .$$

In particular the ordered array of vertices

$$wT(p_0)\ wT(p_1) \cdots wT(p_q)$$

defines a cell of the complex $K(P)$ which we shall denote by τT. Clearly we get a chain transformation

$$\tau : S_V(P) \to K(P)$$

such that

$$\tau\beta_V = 1.$$

It remains to show that

$$\beta_V\tau \simeq 1.$$

Given T, s, and s_1 as before we consider the singular simplex $T' = (\beta_V\tau)T$. We may assume that T' is given as a continuous mapping

$$T':s \to P.$$

Since $T'(s) \subset s_1$ and $T(s) \subset V(s_1)$ it follows that for each point $x \in s$ the points $T(x)$ and $T'(x)$ are both in the same simplex of P; this simplex may of course depend upon x.

We define a singular prism

$$R_T:s \times I \to P$$

as follows: $R_T(x, t)$, for $x \in s$ and $t \in I$, is the point dividing the interval from $T(x)$ to $T'(x)$ in the ratio $t:1 - t$.

We verify that

if $T_1 \equiv T_2$ then $R_{T_1} \equiv R_{T_2}$,

$R_{T^{(i)}} = R_T^{(i)}$,

$R_T(0) = T$, $R_T(1) = (\beta_V\tau)T$.

Consequently if, following §16, we define

$$D(T) = c(R_T)$$

we get by (16.2)

$$\partial D(T) = \beta_{VT}T - T - D(\partial T)$$

which proves that $\beta_{VT} \simeq 1$.

The proof that (20.1) also is a chain equivalence requires further preparation.

21. Barycentric subdivision

Let s be a geometric simplex and let $\sigma(s)$ be its first barycentric subdivision. We shall consider the complexes $K(s)$ and $K[\sigma(s)]$. For each integral chain c^q of $K(s)$ we shall define an integral chain $\sigma(c^q)$ of $K[\sigma(s)]$ subject to the following conditions: Each 0-chain c^0 in $K(s)$ may also be regarded as a chain in $K[\sigma(s)]$, we assume that

$$\sigma(c^0) = c^0. \tag{21.1}$$

For each q-cell α^q of $K(s)$, we denote by $| \alpha^q |$ the smallest simplex of s containing α^q. The polyhedron $\sigma(| \alpha^q |)$ is a subpolyhedron of $\sigma(s)$. We assume that

$$\sigma(\alpha^q) \subset \sigma(| \alpha^q |). \tag{21.2}$$

$$\sigma(c_1^q + c_2^q) = \sigma(c_1^q) + \sigma(c_2^q) \tag{21.3}$$

$$\sigma(\partial c^q) = \partial\sigma(c^q). \tag{21.4}$$

We proceed by induction. Suppose that $\sigma(c^q)$ is defined for $q < n$ as satisfies conditions (21.1) – (21.4). In view of (21.3) it is sufficient to define $\sigma(\alpha^n)$ for each n-cell σ^n of $K(s)$. Let $b = b(| \alpha^n |)$ be the barycenter of $| \alpha^n |$. The subdivision $\sigma(| \alpha^n |)$ of $| \alpha^n |$ may be regarded as a cone with b as vertex, since by (21.2) we have $\sigma(\partial\alpha^n) \subset \sigma(| \alpha^n |)$ the construction of §10 defines a chain $b\sigma(\partial\alpha^n)$ in $K[\sigma(| \alpha^n |)]$ which is a subcomplex of $K[\sigma(s)]$. We define

$$\sigma(\alpha^n) = b\sigma(\partial\alpha^n).$$

Clearly (21.2) is satisfied. Using (10.1) and (21.4) we get

$$\partial\sigma(\alpha^n) = \partial[b\sigma(\partial\alpha^n)] = \sigma(\partial\alpha^n) - b\partial\sigma(\partial\alpha^n) = \sigma(\partial\alpha^n) - b\sigma(\partial\partial\alpha^n) = \sigma(\partial\alpha^n).$$

The chain transformation

$$\sigma: K(s) \to K[\sigma(s)] \tag{21.5}$$

thus obtained has the following important consistency property: if s' is a face of s, then if we regard $K(s')$ as a subcomplex of $K(s)$ and $K[\sigma(s')]$ as a subcomplex of $K[\sigma(s)]$, then the chain transformation

$$\sigma: K(s') \to K[\sigma(s')]$$

agrees with (21.5).

Given a geometric simplex x it will be convenient to consider a simplex $\rho(s)$ whose vertices are the vertices of s and the barycenters of all the faces of all dimensions of s, including the barycenter of s itself. $\rho(s)$ is a simplex of dimension $2(2^q - 1)$ where q is the dimension of s, and may be constructed in a euclidean space of sufficiently high dimension. The complexes $K(s)$ and $K[\sigma(s)]$ may in a natural fashion be regarded as subcomplexes of $K[\rho(s)]$.

We shall now define for each integral chain c^q in $K(s)$ an integral chain $\rho(c^q)$ in $K[\rho(s)]$ subject to the following conditions

$$\rho(c^0) = 0 \tag{21.6}$$

$$\rho(c_1^q + c_2^q) = \rho(c_1^q) + \rho(c_2^q) \tag{21.7}$$

$$\partial\rho(c^q) = \sigma c^q - c^q - \rho(\partial c^q). \tag{21.8}$$

We assume that $\rho(c^q)$ has already been defined for $q < n$. In view of (21.7) it is sufficient to define $\rho(\alpha^n)$ for any n-cell α^n of $K(s)$. As before let $b = b(|\alpha^n|)$ be the barycenter of $|\alpha^n|$. Since b is a vertex of $\rho(s)$ we define as in §10

$$\rho(\alpha^n) = b[\sigma\alpha^n - \alpha^n - \rho(\partial\alpha^n)].$$

Using (10.1) and (21.8) we get

$$\begin{aligned}
\partial\rho(\alpha^n) &= \partial b[\sigma\alpha^n - \alpha^n - \rho(\partial\alpha^n)] \\
&= \sigma\alpha^n - \alpha^n - \rho(\partial\alpha^n) - b[\partial\sigma\alpha^n - \partial\alpha^n - \partial\rho(\partial\alpha^n)] \\
&= \sigma\alpha^n - \alpha^n - \rho(\partial\alpha^n) - b[\partial\sigma\alpha^n - \partial\alpha^n - \partial\sigma\alpha^n + \partial\alpha^n - \rho(\partial\partial\alpha^n)] \\
&= \sigma\alpha^n - \alpha^n - \rho(\partial\alpha^n).
\end{aligned}$$

This completes the definition of $\rho(c^q)$. The function ρ has a similar consistency property as σ. If s' is a subsimplex of s then $\rho(s')$ may be regarded as a subsimplex of $\rho(s)$ and $K[\rho(s')]$ may be regarded as subcomplex of $K[\rho(s)]$. Each chain c^q of $K(s')$ may also be regarded as a chain of $K(s)$. The chain $\rho(c^q)$ will then be the same whether regarded in $K[\rho(s')]$ or in $K[\rho(s)]$.

22. Subdivision of singular simplexes

Let

$$T: s \to X, \qquad s = \langle p_0 \cdots p^q \rangle$$

be a singular q-simplex, and let α^q be the q-cell of $K(s)$ defined by

$$\alpha^q = p_0 \cdots p_q.$$

The chain $\sigma(\alpha^q)$ is then a well defined chain of $K[\sigma(s)]$ and using the notation of §16 we get a singular chain

$$\sigma(T) = (\sigma(s), \sigma(\alpha^q), T)$$

of X. It is easy to verify that this way we get a chain transformation

$$\sigma: S(X) \to S(X).$$

We now consider the simplex $\rho(s)$. Since each vertex of $\rho(s)$ is a point of s we have a linear mapping

$$R:\rho(s) \to s$$

which gives a mapping

$$TR:\rho(s) \to X.$$

We define as singular $(q+1)$-chain in X by setting

$$\rho(T) = (\rho(s), \rho(\alpha^q), TR).$$

From (21.8) we get

(22.1) $$\partial\rho(T) = \sigma(T) - T - \rho(\partial T)$$

which shows that the chain transformation σ is chain homotopic to 1.

The chain transformation σ may be iterated by setting

$$\sigma^0(T) = T, \qquad \sigma^{n+1}(T) = \sigma[\sigma^n(T)].$$

We further remark that both chains $\sigma(T)$ and $\rho(T)$ are contained in the set $T(s)$. Consequently for $X = P$ we get

(22.2) *If T is a singular simplex in $S_V(P)$ then $\sigma(T)$ and $\rho(T)$ are both in $S_V(P)$.*

23. Completion of the proof

Given $T \epsilon S(P)$ we denote by $n(T)$ the smallest integer n such that $\sigma^n(T)$ is a chain in $S_V(P)$. Such an integer exists because of Lemma 19.1. It is clear that if $T^{(i)}$ is a face of T then $n(T^{(i)}) \leqq n(T)$. Obviously T is in $S_V(P)$ if and only if $n(T) = 0$.

We define for T in $S(P)$

(23.1) $$P(T) = \sum_{j=0}^{n(T)-1} \rho\sigma^j(T).$$

Clearly $P(T) = 0$ for T in $S_V(P)$.

Let $T^{(i)}$ be the face of T opposite to the i^{th} vertex so that $\partial T = \sum_{i=0}^{q} (-1)^i T^{(i)}$ where q is the dimension of T. We shall prove that

(23.2) $$\partial P(T) = \sigma^{n(T)}(T) - T - \sum_{i=0}^{q} (-1)^i \sum_{j=0}^{n(T)-1} \rho\sigma^j(T^{(i)})$$

(23.3) $$P(\partial T) = \sum_{i=0}^{q} (-1)^i \sum_{j=0}^{n(T^{(i)})-1} \rho\sigma^j(T^{(i)}).$$

In fact, using (23.1) and (22.1) we have

$$\partial P(T) = \sum_{j=0}^{n(T)-1} \partial\rho\sigma^j(T) = \sum_{j=0}^{n(T)-1} [\sigma^{j+1}(T) - \sigma^j(T) - \rho\sigma^j(\partial T)]$$

$$= \sigma^{n(T)}(T) - T - \sum_{j=0}^{n(T)-1} \rho\sigma^j \sum_{j=0}^{q} (-1)^i T^{(i)}.$$

This proves (23.2). On the other hand using (23.1) alone we have

$$P(\partial T) = \sum_{i=0}^{q} (-1)^i P(T^{(i)}) = \sum_{j=0}^{q} (-1)^i \sum_{j=0}^{n(T^{(i)})-1} \rho\sigma^j(T^{(i)})$$

which proves (23.3).

From (23.2) and (23.3) we get

$$T + \partial P(T) + P(\partial T) = \sigma^{n(T)}(T) - \sum_{i=0}^{q} (-1)^i \sum_{j=n(T^{(i)})}^{n(T)-1} \rho\sigma^j(T^{(i)}).$$

This implies that

(23.4) *For each q-dimensional singular complex T in $S(P)$ the chain*

$$T + \partial P(T) + P(\partial T)$$

is in $S_V(T)$.

We define

(23.5) $$\pi(T) = T + \partial P(T) + P(\partial T)$$

for each T in $S(P)$. It follows from §5 and from (23.4) that

(23.6) $$\pi: S(P) \to S_V(P)$$

is a chain transformation. Let

(23.7) $$\epsilon: S_V(P) \to S(P)$$

be the identity chain transformation: $\epsilon(T) = T$ for $T \epsilon S_V(P)$. For $T \epsilon S_V(P)$ we have $P(T) = 0$ and also $P(\partial T) = 0$ so that $\pi\epsilon(T) = \pi(T) = T$. Hence

(23.8) $$\pi\epsilon = 1$$

Further, since $\epsilon\pi(T) = \pi(T)$, it follows from (23.5) that

$$\partial P(T) = \epsilon\pi(T) - T - P(\partial T)$$

which shows that

(23.9) $$\epsilon\pi \simeq 1.$$

Formulae (23.8) and (23.9) show the ϵ is a chain equivalence. This completes the proof of Theorem 14.1.

CHAPTER V

PRODUCTS[20]

24. Augmentable complexes

The purpose of introducing the definition of an augmentable complex here is to secure that the products that will subsequently be studied have a unit.

[20] General reference: S. Lefschetz, *Algebraic Topology*, Chapters III and IV; H. Whitney, *On products in a complex*, Ann. of Math., 39 (1938), pp. 397–432.

A closure finite abstract complex K will be called *augmentable* provided the following two properties hold.

(24.1) *All the cells of K have dimension $\geqq 0$ and there are cells of dimension* 0.

(24.2) *The integral* 0*-cochain f_0 defined by $f_0(\sigma^0) = 1$ for all σ^0 in K is a cocycle.*

An augmentable complex can be *augmented* by the addition of a single (-1)-dimensional cell σ^{-1} with incidence numbers $[\sigma^0, \sigma^{-1}] = 1$ for all 0-cells σ^0. Then f_0 becomes a coboundary.

Let K be augmentable. Given a 0-chain

$$c^0 = \sum g_i\sigma_i^0 \in C^0(K, G)$$

we define the *index* $I(c^0)$ of c^0 as

$$I(c^0) = \sum g_i \in G$$

and verify at once that

(24.3) $$I(c^0) = KI(f_0, c^0).$$

We verify by inspection that

(24.4) *The complexes $k(P)$, $K(P)$ and $S(X)$ are augmentable.*

As we have remarked earlier these complexes will be left unaugmented.

25. The cup product

The complexes that are studied in this paper do not all fall in the class that Whitney calls "complexes admitting a product theory." Particularly our complex $S(X)$ is not such a complex. As a result Whitney's theory establishing in a general fashion the existence and the uniqueness of the products cannot be used here. The products will have to be defined individually for each class of complexes $k(P)$, $K(P)$, $S(X)$, and others that will appear later. Nevertheless all these products have enough common features to permit a uniform axiomatic treatment that will be developed in the sequel.

An augmentable closure finite abstract complex K will be called a *complex with products* if a rule is given which with any three groups G_1, G_2, G such that G_1 and G_2 are paired to G (see §7), and with any two cochains

$$f_1^p \in C_p(K, G_1) \qquad f_2^q \in C_q(K, G_2)$$

associates a third cochain, called the *cup product of f_1^p and f_2^q*

$$f_1^p \cup f_2^q \in C_{p+q}(K, G)$$

subject to the following five axioms:

($\cup$ 1) $\quad f_1^p \cup f_2^q$ is additive and continuous in each variable.

This axiom merely states that under the cup product as multiplication the groups $C_p(K, G_1)$ and $C_q(K, G_2)$ are paired to the group $C_{p+q}(K, G)$.

(∪ 2) Associativity: $f_1^p \cup (f_2^q \cup f_3^r) = (f_1^p \cup f_2^q) \cup f_3^r$.

More precisely if G_1, G_2, G_3, G_{12}, G_{23} and G are six groups with pairings

$$g_1g_2 \in G_{12} \qquad g_2g_3 \in G_{23}$$

$$g_1g_{23} \in G \qquad g_{12}g_3 \in G$$

such that

$$g_1(g_2g_3) = (g_1g_2)g_3$$

then (∪ 2) holds for

$$f_1^p \in C_p(K, G_1), \qquad f_2^q \in C_q(K, G_2), \qquad f_3^r \in C^r(K, G_3).$$

(∪ 3) $$f \cup f_0 = f$$

(∪ 4) $$f_0 \cup f = f.$$

In these two axioms f_0 is the integral 0-cocycle described in the previous section. The group I of integers is considered paired with any group G in the obvious fashion.

(∪ 5) $$\delta(f_1^p \cup f_2^q) = \delta f_1^p \cup f_2^q + (-1)^p f_1^p \cup \delta f_2^q$$

It follows from (∪ 5) that

cocycle ∪ cocycle = cocycle

cocycle ∪ coboundary = coboundary

coboundary ∪ cocycle = coboundary;

consequently the cup product is defined for cohomology classes and gives a pairing of the cohomology groups $H_p(K, G_1)$ and $H_q(K, G_2)$ to the group $H_{p+q}(K, G)$.

We can now see why we do not consider the complexes K augmented. Let f be any cocycle. If K were augmented than f_0 would be a coboundary and consequently $f = f \cup f_0$ would also be a coboundary.

26. The cap product

The cap product $f^q \cap c^{p+q}$ defined for a cochain f^q and a chain c^{p+q} can be derived from the cup product using characters and the Kronecker index.

Given a pairing of the groups G_1 and G_2 to G with G_2 and G discrete, we define a new pairing of the groups Char G and G_1 to the group Char G_2 by setting

$$(\chi g_1)(g_2) = \chi(g_1g_2)$$

for $\chi \in$ Char G, $g_1 \in G_1$ and $g_2 \in G_2$. Clearly χg so defined is a character of G_2 and has all the properties of a pairing.

Let now

$$f^q \in C_q(K, G_1), \qquad c^{p+q} \in C^{p+q}(K, G_2)$$

with the groups G_1 and G_2 paired to G as above. Consider an arbitrary cochain $g^p \in C_p(K, \text{Char } G)$. Since G and G_1 are paired to Char G_2 the cochain

$$g^p \cup f^q \in C_{p+q}(K, \text{Char } G_2)$$

is defined. Since Char G_2 and G_2 are paired to the group R_1 of reals reduced mod 1 the Kronecker index

$$KI(g^p \cup f^q, c^{p+q})$$

is defined, and for fixed f^q and c^{p+q} gives a character of the group $C_p(K, \text{Char } G)$. Since the character group of $C_p(K, \text{Char } G)$ is the group $C^p(K, G)$ (cf. (7.3)) there is a unique chain

$$f^q \cap c^{p+q} \in C^p(K, G)$$

such that

$$KI(g^p \cup f^q, c^{p+q}) = KI(g^p, f^q \cap c^{p+q}) \tag{26.1}$$

for all cochains $g^p \in C_p(K, \text{Char } G)$.

Axioms ($\cup$ 1) – ($\cup$ 5) of the cup product translate into the following properties of the cap product

($\cap$ 1) $f^q \cap c^{p+q}$ is additive and continuous in both variables

($\cap$ 2) $$f^q \cap (f^r \cup c^{p+q+r}) = (f^q \cap f^r) \cup c^{p+q+r}$$

($\cap$ 3) $$f_0 \cap c^p = c^p$$

($\cap$ 4) $$I(f^q \cap c^q) = KI(f^q, c^q)$$

($\cap$ 5) $$\partial(f^q \cap c^{p+q}) = (-1)^p \delta f^q \cap c^{p+q} + f^q \cap \partial c^{p+q}.$$

We will only prove ($\cap$ 5). Let g^{p-1} be any cochain of $C_{p-1}(K, \text{Char } G)$. From (7.1), ($\cup$ 5) and (26.1) we get

$$\begin{aligned} KI(g^{p-1}, f^q \cap \partial c^{p+q}) &= KI(g^{p-1} \cup f^q, \partial c^{p+q}) = KI(\delta(g^{p-1} \cup f^q), c^{p+q}) \\ &= KI(\delta g^{p-1} \cup f^q, c^{p+q}) + KI((-1)^{p-1} g^{p-1} \cup \delta f^q, c^{p+q}) \\ &= KI(\delta g^{p-1}, f^q \cap c^{p+q}) + KI(g^{p-1}, (-1)^{p-1} \delta f^q \cap c^{p+q}) \\ &= KI(g^{p-1}, \partial(f^q \cap c^{p+q})) + KI(g^{p-1}, -(-1)^p \delta f^q \cap c^{p+q}). \end{aligned}$$

This implies

$$KI(g^{p-1}, \partial(f^q \cap c^{p+q})) = KI(g^{p-1}, (-1)^p \delta f^q \cup c^{p+q} + f^q \cap \partial c^{p+q}).$$

Since this is true for every cochain $g^{p-1} \in C_{p-1}(K, \text{Char } G)$ formula ($\cap$ 5) follows.

From ($\cap$ 5) we deduce that

cocycle $\cap$ cycle = cycle

cocycle $\cap$ boundary = boundary

coboundary $\cap$ cycle = boundary;

consequently the cap product is defined for cohomology and homology classes and gives a pairing of the groups $H_q(K, G_1)$ and $H^{p+q}(K, G_2)$ to the group $H^q(K, G)$.

27. Chain transformations preserving products

Let K_1 and K_2 be two complexes with products, and let

$$\tau: K_1 \to K_2$$

be a chain transformation. We shall say that τ *preserves the products* provided

$$\tau^*(f_1^p \cup f_2^q) = \tau^* f_1^p \cup \tau^* f_2^q \tag{27.1}$$

for $f_1^p \in C^p(K_2, G_1)$ and $f_2^p \in C^p(K_2, G_2)$, with G_1 and G_2 paired to some group G.

If τ preserves the products then for the cap product we have the formula

$$\tau(\tau^* f^q \cap c^{p+q}) = f^q \cap \tau c^{p+q} \tag{27.2}$$

for $f^q \in C_q(K_2, G_1)$, $c^{p+q} \in C^{p+q}(K_1, G_2)$ with G_1 and G_2 paired to G. In order to prove (27.2) we consider an arbitrary cochain $g^p \in C_p(K_2, \text{Char } G)$. Using (7.2), (26.1) and (27.1) we get

$$\begin{aligned} KI(g^p, \tau(\tau^* f^q \cup c^{p+q})) &= KI(\tau^* g^p, \tau^* f^q \cap c^{p+q}) = KI(\tau^* g^p \cup \tau^* f^p, c^{p+q}) \\ &= KI(\tau^*(g^p \cup f^q), c^{p+q}) = KI(g^p \cup f^q, \tau c^{p+q}) = KI(g^p, f^q \cap \tau c^{p+q}). \end{aligned}$$

Since this holds for every g^p, (27.2) follows.

If formula (27.1) is not necessarily true for all cochains f_1^p, f_2^q but is true for any two cohomology classes, then we say that τ *weakly* preserves the products. Formula (27.2) is then still valid for cohomology and homology classes.

If τ weakly preserves the product, so does every chain transformation chain homotopic with τ. If τ and ρ are an equivalence pair and τ weakly preserves the products, then the same is true for ρ.

28. Products in $S(X)$, $K(P)$ and $k(P)$

As before we assume that the groups G_1 and G_2 are paired to G.

We first define the cup product in the complex $S(X)$. Let then two cochains

$$f_1^p \in C_p(X, G_1) \quad \text{and} \quad f_2^q \in C_q(X, G_2)$$

be given. Consider a singular $(p + q)$-simplex

$$T: s \to X, \qquad s = < p_0 \cdots p_{p+q} >$$

and define

$$\begin{aligned} {}_pT &= T \mid s_1 \to X \quad \text{where} \quad s_1 = < p_0 \cdots p_p > \\ T_q &= T \mid s_2 \to X \quad \text{where} \quad s_2 = < p_p \cdots p_{p+q} >. \end{aligned}$$

Clearly T_1 is a singular p-simplex and T_2 is a singular q-simplex in X. We define

$$(f_1^p \cup f_2^q)(T) = f_1^p({}_pT) f_2^q(T_q). \tag{28.1}$$

Clearly $f_1^p \cup f_2^q$ is a well defined $(p + q)$-cochain in X with coefficients in G. Axioms $(\cup\ 1)$–$(\cup\ 5)$ can be verified by straightforward computation, and thus $S(X)$ is a complex with products.

As shown in §26 the definition of the cup product automatically implies a definition of the cap product. In this case the cap product can easily be seen to be

$$f^q \cap (g_2 T) = [f^q(T_q)g_2]_p T \tag{28.2}$$

with T, ${}_pT$ and T_q related as above.

Next we turn to the complex $K(P)$. In this case the definition of $f_1^p \cup f_2^q$ is even simpler. It reads

$$(f_1^p \cup f_2^q)(v_0 \cdots v_{p+q}) = f_1^p(v_0 \cdots v_p)f_2^q(v_p \cdots v_{p+q}). \tag{28.3}$$

The verification of axioms $(\cup\ 1) - (\cup\ 5)$ is immediate. The related definition of the cap product is

$$f^q \cap g_2 v_0 \cdots v_{p+q} = [f^q(v_p \cdots v_{p+q})g_2]v_0 \cdots v_p\,. \tag{28.4}$$

If we now return to the definition of the basic chain transformation (§14)

$$\beta : K(P) \to S(P)$$

it becomes evident that

(28.5) *β preserves the products.*

Since β was proved to be a chain equivalence it follows that the complexes $K(P)$ and $S(P)$ not only have isomorphic homology and cohomology groups but also isomorphic product theories. This proves the topological invariance of the products in $K(P)$.

In order to introduce the products in the complex $k(P)$ we choose a definite order for all the vertices of P. Each $(p + q)$-cell σ^{p+q} of $k(P)$ can then uniquely be written as (cf. §9 and 11)

$$\sigma^{p+q} = \pm [v_0 \cdots v_{p+q}], \qquad v_0 < \cdots < v_{p+q}$$

We define

$$(f_1^p \cup f_2^q)([v_0 \cdots v_{p+q}]) = f_1^p([v_0 \cdots v_p])f_2^q([v_p \cdots v_{p+q}]) \tag{28.5}$$

Axioms $(\cup\ 1)$–$(\cup\ 5)$ are easy to verify. The related definition of the cap product reads

$$f^q \cap g_2[v_0 \cdots v_{p+q}] = f^q([v_p \cdots v_q])g_2[v_0 \cdots v_p] \tag{28.6}$$

We recall now that in order to prove that the chain transformation

$$\alpha : K(P) \to k(P)$$

defined in §9 was a chain equivalence we have constructed (§11) a chain transformation

$$\bar{\alpha} : k(P) \to K(P)$$

such that α and $\bar{\alpha}$ were an equivalence pair. In order to define $\bar{\alpha}$ we have assumed that a definite order of the vertices of P was chosen. If we now agree to use the same order of vertices of P for the definition of $\bar{\alpha}$ as for the definition of the product in $k(P)$ it becomes clear by inspection that

(28.7) *$\bar{\alpha}$ preserves the products.*

By §27 this implies that

(28.8) *α weakly preserves the products.*

This shows that as far as cohomology and cohomology classes are concerned the products in $k(P)$ do not depend upon the choice of the order of the vertices of P, initially used in the definition of $f_1^p \cup f_2^q$. It also proves the topological invariance of the products in $k(P)$.

If X and Y are topological spaces and

$$\varphi: X \to Y$$

a continuous mapping, then it is easy to see that the chain transformation

$$\varphi: S(X) \to S(Y)$$

preserves the products.

29. Commutativity of the products

Let the (topologized) groups G_1 and G_2 be paired to G. For two cohomology classes

$$f_1^p \in H_p(X, G_1) \quad \text{and} \quad f_2^q \in H_q(X, G_2)$$

the cohomology class

$$f_1^p \cup f_2^q \in H_{p+q}(X, G)$$

is defined.

We also define a pairing of G_2 and G_1 to G by setting

$$g_2 g_1 = g_1 g_2$$

and consequently get the cohomology class

$$f_2^q \cup f_1^p \in H^p_{+q}(X, G)$$

THEOREM 29.1.

$$f_1^p \cup f_2^q = (-1)^{pq} f_2^q \cup f_1^p.$$

We emphasize that the theorem is valid only for cohomology classes and not for individual cocycles.

Before proceeding with the proof we define a chain transformation

$$\rho: S(X) \to S(X)$$

as follows. Let $T: s \to X$ be a singular simplex in X where $s = < p_0 \cdots p_q >$. Consider the simplex $\bar{s}$ obtained from s by reversing the order of vertices: $\bar{s} =$

$< p_q \cdots p_0 >$. Setting $\overline{T}(x) = T(x)$ we get a new singular simplex $\overline{T}:\bar{s} \to X$ of $S(X)$. We define

$$\rho(T) = (-1)^{\frac{1}{2}q(q+1)}\,\overline{T}.$$

We verify by a straightforward computation that $\rho\partial = \partial\rho$ so that ρ is a chain transformation.

LEMMA 29.2. $\rho \simeq 1$.

PROOF. Let s be a geometric simplex. We define a chain transformation

$$\rho: K(s) \to K(s)$$

by setting for each q-cell $\sigma^q = v_0 \cdots v_q$ of $K(s)$

$$\rho(\sigma^q) = (-1)^{\frac{1}{2}q(q+1)} v_q \cdots v_0.$$

We verify easily that $\rho\partial = \partial\rho$. Next we define homomorphisms

$$\tau: C^q(K(s)) \to C^{q+1}(K(s))$$

by induction, using the notations of §10, as follows

$$\tau(c^0) = 0$$

$$\tau(v_0 \cdots v_q) = v_0[v_0 \cdots v_q - \rho(v_0 \cdots v_q) - \tau(\partial v_0 \cdots v_q)].$$

From 10.1 we get

$$\partial\tau(v_0 \cdots v_q) = v_0 \cdots v_q - \rho(v_0 \cdots v_q) - \tau(\partial v_0 \cdots v_q)$$
$$-v_0[\partial v_0 \cdots v_q - \rho(\partial v_0 \cdots v_q) - \partial\tau(\partial v_0 \cdots v_q)]$$

hence by induction it follows that

$$\partial\tau\sigma^q = \sigma^q - \rho\sigma^q - \tau\partial\sigma^q. \tag{29.1}$$

The chain transformation ρ and the homotopy operator τ have the important consistency property as described in §21.

Now let

$$T: s \to X \qquad s = < p_0 \cdots p_q >$$

be a singular q-simplex. We consider the q-cell $\sigma^q = p_0 \cdots p_q$ of the complex $K(s)$ and using the notations of §16 we define

$$\tau T = (s, \sigma^q, T)$$

τT is a well defined singular chain of X and from (29.1) and (16.2) we get

$$\partial\tau T = T - \rho T - \tau\partial T.$$

This shows that $\rho \simeq 1$ q.e.d.

We now proceed with the proof of Theorem 29.1. Let two cochains

$$f_1^p \in C_p(X, G_1) \qquad f_2^q \in C_q(X, G_2)$$

be given with G_1 and G_2 paired to G, and let T be a singular simplex of dimension $p + q$.

From the definition of $\overline{T}$ we have

$$_q(\overline{T}) = \overline{(T_q)}, \qquad (\overline{T})_p = \overline{({}_pT)}.$$

Hence assuming that $g_2g_1 = g_1g_2$

$$f_2^q \cup f_1^p(\overline{T}) = f_2^q[_q(\overline{T})]f_1^p[(\overline{T})_p] = f_1^p[\overline{({}_pT)}]f_2^q[\overline{(T_q)}]$$

and using the definition of ρ we get

$$(-1)^{\frac{1}{2}(p+q)(p+q+1)}(f_2^q \cup f_1^p)(\rho T) = (-1)^{\frac{1}{2}p(p+1)}(-1)^{\frac{1}{2}q(q+1)}f_1^p(\rho_pT)f_2^q(\rho T_q)$$

and after simplification

$$(-1)^{pq}(f_2^q \cup f_1^p)(\rho T) = f_1^p(\rho_pT)f_2^q(\rho T_q).$$

But since

$$f_2^q \cup f_1^p(\rho T) = \rho^*(f_2^q \cup f_1^p)(T)$$

$$f_1^p(\rho_pT) = \rho^*f_1^p({}_pT), \qquad f_2^q(\rho T_q) = \rho^*f_2^q(T_q)$$

we obtain

$$(-1)^{pq}\rho^*(f_2^q \cup f_1^p)(T) = \rho^*f_1^p({}_pT)\rho^*f_2^q(T_q) = (\rho^*f_1^p \cup \rho^*f_2^q)(T)$$

and therefore

$$\rho^*f_1^p \cup \rho^*f_2^q = (-1)^{pq}\rho^*(f_2^q \cup f_1^p).$$

Since $\rho \simeq 1$ we have $\rho^*f = f$ for every cohomology class f. Hence for cohomology classes we have

$$f_1^p \cup f_2^q = (-1)^{pq}f_2^p \cup f_1^q.$$

Having proved the commutativity property of the product in $S(X)$, a similar formula for the complexes $K(P)$ and $k(P)$ follows using the chain transformations β and α.

CHAPTER VI

RELATIONS WITH HOMOTOPY GROUPS

30. The complexes $S_n(X)$

Let X be a topological space and x_0 a point of X. We shall use x_0 as a base point for the construction of the homotopy groups[21] $\pi_n(X)$ of X. The real interest of the homotopy groups is only for arcwise connected spaces. If X is not arcwise connected then $\pi_n(X)$ coincides with $\pi_n(X_0)$ where X_0 is the arc-component of x_0 in X. For convenience the fact that X is arcwise connected will be recorded as $\pi_0(X) = 0$.

For each positive integer n we consider the subcomplex $S_n(X)$ of $S(X)$ defined as follows. A singular q-simplex $T: s \rightarrow X$ is in $S_n(X)$ if and only if all the faces

---

[21] W. Hurewicz, Proc. Akad. Amsterdam, 38 (1935), pp. 112–119 and pp. 521–528; 39 (1936), pp. 117–126 and pp. 215–224.

of s of dimension $< n$ are mapped by T into x_0. Thus $S_1(X)$ consists of all the singular simplices whose vertices are mapped into x_0. We also define $S_0(X) = S(X)$. We thus obtain a descending sequence of closed subcomplexes

$$S(X) = S_0(X) \supset S_1(X) \supset S_2(X) \supset \cdots.$$

The identity mapping leads to chain transformations

$$\eta_{n,m}: S_n(X) \to S_m(X) \qquad n \geqq m.$$

Instead of $\eta_{n,0}$ we shall also write η_n, so that

$$\eta_n: S_n(X) \to S(X).$$

We notice that

$$\eta_{m,l}\,\eta_{n,m} = \eta_{n,l}.$$

Let X and Y be two topological spaces with base points x_0 and y_0 respectively. Given a continuous mapping

$$\varphi: X \to Y$$

such that $\varphi(x_0) = y_0$, the chain transformation

$$\varphi: S(X) \to S(Y)$$

defined in §15 will carry the subcomplexes $S_n(X)$ into $S_n(Y)$

$$\varphi: S_n(X) \to S_n(Y)$$

THEOREM 30.1. *If the continuous mappings $\varphi: X \to Y$ and $\psi: X \to Y$ such that $\varphi(x_0) = y_0$ and $\psi(x_0) = y_0$ are homotopic relative to x_0 (i.e. are homotopic with $x_0 \to y_0$ throughout the homotopy) then the induced chain transformations $\varphi: S_n(X) \to S_n(Y)$ and $\psi: S_n(X) \to S_n(Y)$ are chain homotopic.*

The proof is identical with that of Theorem 17.1.

31. A reduction theorem[8]

The main reason for the usefulness of the complexes $S_n(X)$ is the following

THEOREM 31.1. *If $\pi_n(X) = 0$ then the chain transformation*

$$\eta_{n+1,n}: S_{n+1}(X) \to S_n(X)$$

is a chain equivalence.

In particular we have

COROLLARY 31.2. *If X is arcwise connected then*

$$\eta_1: S_1(X) \to S(X)$$

is a chain equivalence.

COROLLARY 31.3. *If $\pi_i(X) = 0$ for $i < n$ then*

$$\eta_n: S_n(X) \to S(X)$$

is a chain equivalence.

Consequently for spaces X such that $\pi_i(X) = 0$ for $i < n$ the homology and cohomology theory of X can be studied by means of singular simplices $T:s \to X$ which map all the faces of s of dimension $< n$ into the base point x_0.

PROOF OF THEOREM 31.1. Let $\pi_n(X) = 0$. We shall define for each $T \in S_n(X)$

$$T:s \to X$$

a singular prism

$$R_T:s \times I \to X$$

subject to the following conditions

(31.1) If $T_1 \equiv T_2$ then $R_{T_1} \equiv R_{T_2}$

(31.2) $R_T(i) = R_T^{(i)}$

(31.3) $R_T(0) = T$

(31.4) $R_T(1)$ is in $S_{n+1}(X)$

(31.5) If $T \in S_{n+1}(X)$ then $R_T(x, t) = T(x)$.

Let q be the dimension of T. If $q < n$ then $T \in S_n(X)$ implies $T:s \to x_0$ and therefore $T \in S_{n+1}(X)$ so that R_T is defined by (31.5). For $q = n$ the mapping $T:s \to X$ sends the boundary $B(s)$ of s into the point x_0. From conditions (31.2), (31.3) and (31.4) it follows that the mapping R_T is already defined on the boundary $B(s \times I)$ of the $(n + 1)$-dimensional prism $s \times I$. Since $B(s \times I)$ is homeomorphic to an n-sphere and $\pi_n(X) = 0$ it follows that R_T can be extended through the interior of $s \times I$.

From this point we proceed by induction. We assume that R_T has already been defined for all singular simplices T of dimension $< q$ in $S_n(X)$, that (31.1) – (31.5) hold. Let $T:s \to X$ be a q-simplex in $S_n(X)$. It follows from (31.2) and (31.3) that the mapping R_T is already given on the subset

$$L = (s \times 0) \cup (B(s) \times I)$$

of $s \times I$. Since the set L is an absolute retract (in fact L is homeomorphic to a q-simplex) it follows that R_T can be extended throughout $s \times I$. It is clear now that the definition of R_T can be so continued as to keep conditions (31.1)–(31.5) satisfied.

Now we define for each singular q-simplex T in $S_n(X)$

$$\pi(T) = R_T(1),$$

$$P(T) = c(R_T).$$

It follows from (16.2) that

$$\partial P(T) = R_T(1) - R_T(0) - \sum_{i=0}^{q} (-1)^i \, c(R_T{}^{(i)}).$$

Hence using (31.2) and (31.3) we find

(31.6) $$\partial P(T) = \pi(T) - T - P(\partial T).$$

This implies by §5 that π leads to a chain transformation, and from (31.4) we get that

$$\pi: S_n(X) \to S_{n+1}(X).$$

From (31.5) we find that $\pi\eta_{n+1,n}(T) = T$ so that

$$\pi\eta_{n+1,n} = 1.$$

Since $\eta_{n+1,n}\pi(T) = \pi(T)$ it follows from (31.6) that

$$\eta_{n+1,n}\pi \simeq 1$$

Hence $\eta_{n+1,n}$ is a chain equivalence, as desired.

Theorem 31.1 just established lends itself to various generalizations. For instance we could replace the point x_0 by a subset A of X and define a subcomplex $S_{1A}(X)$ consisting of all the singular cells in X whose vertices are in A. An examination of the previous proof shows that if X is arcwise connected then the identity chain transformation mapping $S_{1A}(X)$ into $S(X)$ is a chain equivalence.

In the complexes $S_n(X)$ the cup product can be defined exactly as in the complex $S(X)$. It is then clear that all the chain transformations $\eta_{n,m}$ preserve the products.

32. Comparison of homology and homotopy groups[21, 8]

We shall now turn to investigate the continuous mappings

$$\varphi: S^q \to X$$

of a q-dimensional spherical surface S^q into the space X. We select a point $s_0 \in S^q$ as base point and assume that $\varphi(s_0) = x_0$. The mappings φ are divided into homotopy classes relative to the point s_0, and these homotopy classes are in a well known fashion the elements of the q^{th} homotopy group $\pi_q(X)$ of X.

The integral homology group $H^q(S^q)$ is an infinite cyclic group. In order to realize that, notice that S^q could be regarded as a simplicial polyhedron. We then have three abstract complexes $S(S^q)$, $K(S^q)$ and $k(S^q)$ all of which have isomorphic homology groups, and the group $H^q(k(S^q))$ is known to be cyclic infinite. The sphere S^q will be *oriented* by selecting a basic homology class z_0^q in the group $H^q(S^q) = H^q(S(S^q))$.

Each mapping $\varphi: S^q \to X$ determines a chain transformation $\varphi: S(S^q) \to S(X)$ and therefore $\varphi(z_0^q)$ is a homology class in X. By Theorem 30.1 the homology class does depend upon the choice of φ within a homotopy class. Consequently we get a mapping

$$\nu_q: \pi_q(X) \to H^q(X) \tag{32.1}$$

From the definition of the group operations in $\pi_q(X)$ it follows that ν_q is a homomorphism.

Let $n \leqq q$. Since $\pi_i(S^q) = 0$ for all $i < n$ it follows from corollary 31.3 that the basic homology class z_0^q of $H^q(S^q)$ could have been chosen to be a homology

class of the subcomplex $S_n(S^q)$. Thus $\varphi(z_0^q)$ would be a homology class in the group $H^q(S_n(X))$, and corresponding to (32.1) we get the homomorphisms

$$\nu_{q,n}:\pi_q(X) \to H^q(S_n(X)) \qquad n \leqq q. \tag{32.2}$$

We verify at once that

$$\nu_{q,n} = \eta_{q,n}\nu_{q,q} \tag{32.3}$$

$$\nu_q = \eta_q\nu_{q,q} \tag{32.4}$$

Hence the homomorphism

$$\nu_{q,q}:\pi_q(X) \to H^q(S_q(X)) \tag{32.5}$$

is more important than the others.

THEOREM 32.1. *For $q > 1$ the homomorphism (32.5)*

$$\nu_{q,q}:\pi_q(X) \to H^q(S_q(X))$$

is an isomorphism of $\pi_q(X)$ onto $H^q(S_q(X))$.

PROOF. Each singular q-simplex T in $S_q(X)$ is a mapping $T:s \to X$ of an ordered q-simplex s, such that the boundary $B(s)$ is mapped into x_0. By the definition of $\pi_q(X)$ each such T defines uniquely an element $\pi(T)$ of $\pi_q(X)$. Since $\pi_q(X)$ is abelian for $q > 1$, we may extend π by linearity to be defined for every q-chain in $S_q(X)$. Further, from the definition of addition in $\pi_q(X)$ it follows directly that $\pi(\partial T) = 0$ for each singular $(q + 1)$-simplex in $S_q(X)$. Hence π is zero on every bounding cycle and we get a homomorphism

$$\pi:H^q(S_q(X)) \to \pi_q(X)$$

The proof that π is the inverse of $\nu_{q,q}$ is left to the reader.

THEOREM 32.2. *If $\pi_i(X) = 0$ for all $i < q$ where $q > 1$ then the homomorphism* (32.1)

$$\nu_q:\pi_q(X) \to H^q(X)$$

is an isomorphism of $\pi_q(X)$ onto $H^q(X)$.

In fact by (32.4) we have $\nu_q = \eta_q\nu_{q,q}$. Since $\nu_{q,q}$ is an isomorphism onto by the previous theorem and $\eta_q:H^q(S_q(X)) \to H^q(X)$ is an isomorphism onto by Corollary 31.3, the theorem follows.

THEOREM 32.3. *The homomorphism*

$$\nu_{1,1}:\pi_1(X) \to H^1(S_1(X))$$

maps the factor group of $\pi_1(X)$ by its commutator group, isomorphically onto the group $H^1(S_1(X))$.

PROOF. Let $c_1(X)$ be the commutator group of $\pi_1(X)$. Clearly we may consider $\nu_{1,1}$ as a homomorphism

$$\nu_{1,1}:\pi_1(X)/c_1(X) \to H^1(S_1(X)). \tag{32.6}$$

As in the proof of Theorem 32.1 we define $\pi(T)$ for any singular 1-simplex T in $S_1(X)$. For an arbitrary 1-chain c^1 in $S_1(X)$, the element $\pi(c^1)$ of $\pi_1(X)$ modulo $c_1(X)$ is defined uniquely. We get a mapping

$$\pi: H^1(S_1(X)) \to \pi_1(X)/c_1(X)$$

which again is the inverse of (32.6).

THEOREM 32.4. *If X is arcwise connected then the homomorphism*

$$\nu_1: \pi_1(X) \to H^1(X)$$

maps the factor group of $\pi_1(X)$ by its commutator subgroup isomorphically onto $H^1(X)$.

The proof is the same as for Theorem 32.2.

33. Spherical cycles

A cycle z^q in $S_n(X)$ $(n \leqq q)$ will be called spherical if it is homologous in $S_n(X)$ to a cycle in $S_q(X)$. The corresponding homology classes will also be called spherical and the resulting subgroup of the group $H^q(S_n(X), G)$ will be denoted by $\Sigma^q(S_n(X), G)$. In formulae we have

$$\Sigma^q(S_n(X, G)) = \eta_{q,n}[H^q(S_q(X)), G] \qquad n \leqq q \tag{33.1}$$

$$\Sigma^q(X, G) = \eta_q[H^q(S_q(X)), G] \tag{33.2}$$

$$\Sigma^q(S_q(X), G) = H^q(S_q(X), G) \tag{33.3}$$

$$\Sigma^q(S_n(X), G) = \eta_{m,n}[\Sigma^q(S_m(X), G)] \qquad n < m \leqq q \tag{33.4}$$

By theorems 32.1 and 32.2 the homomorphism $\nu_{q \cdot q}$ maps the homotopy group $\pi_q(X)$ onto $H^q(S_n(X))$. Since $\nu_{q,n} = \nu_{q,n}\nu_{q,q}$ and $\nu_q = \eta_q\nu_{q,q}$ it follows from (33.1) and (33.2) that

$$\Sigma^q(S_n(X)) = \nu_{q,n}[\pi_q(X)] \qquad n \leqq q \tag{33.5}$$

$$\Sigma^q(X) = \nu_q[\pi_q(X)]. \tag{33.6}$$

This shows that the subgroups Σ^q for integral coefficients are the images of the q^{th} homotopy group under the suitable natural homomorphism. This justifies the term spherical used in the definition.

LEMMA 33.7. *A homology class $z^q \in H^q(S_n(X), G)$ is in $\Sigma^q(S_n(X), G)$ if and only if $z^q = \sum_{i=1}^{k} g_i z_i^q$ where $g_i \in G$ and $z_i^q \in \Sigma^q(S_n(X))$.*

PROOF. If $z^q(S_n(X), G)$ then the homology class z^q contains a cycle z_1^q contained in the complex $S_q(X)$. Since $S_q(X)$ contains only one $(q - 1)$-cell it easily follows that every q-cycle in $S_q(X)$ is a linear combination with coefficients in G of integral cycles. This shows that the z^q is of the form $\sum g_i z_i^q$. The converse is obvious.

34. The groups Λ

Since any group G is paired with the group of integers, the Kronecker index establishes a pairing of the groups $H^q(X)$ and $H_q(X, G)$ with values in G. With

respect to this pairing we define the subgroup $\Lambda_q(X, G)$ of $H_q(X, G)$ as follows:

$$\Lambda_q(X, G) = \text{Annihilator of } \Sigma^q(X).$$

In a similar fashion we consider the pairing of $H^q(S_n(X))$ and $H_q(S_n(X), G))$ to G and define

$$\Lambda_q(S_n(X), G)) = \text{Annihilator of } \Sigma^q(S_n(X)), \qquad n \leqq q.$$

The identity chain transformations

$$\eta_{n,m}: S_n(X) \to S_m(X) \qquad n \geqq m$$

induce homomorphisms

$$\overset{*}{\eta}_{n,m}: H_q(S_m(X), G) \to H_q(S_n(X), G)) \qquad m \leqq n \leqq q.$$

We shall examine the behavior of the subgroups $\Lambda_q(S_m(X), G)$ under these homomorphisms.

THEOREM 34.1. *A cohomology class f^q in $H_q(S_m(X), G)$ belongs to $\Lambda_q(S_m(X), G)$ if and only if $\overset{*}{\eta}_{n,m} f^q$ is in $\Lambda_q(S_n(X), G)$. In a formula*

$$\Lambda_q(S_m(X), G) = \overset{*-1}{\eta}_{n,m}[\Lambda_q(S_n(X), G)], \qquad m \leqq n \leqq q.$$

PROOF. By definition $f^q \in \Lambda_q(S_m(X), G))$ if and only if $KI(f^q, z^q) = 0$ for all $z^q \in \Sigma^q(S_m(X))$. By lemma 33.2 we have $\Sigma^q(S_m(X)) = \eta_{n,m}[\Sigma^q(S_n(X))]$, consequently $f^q \in \Lambda_q(S_m(X), G)$ if and only if $KI(f^q, \eta_{n,m} z^q) = 0$ for all $z^q \in \Sigma^q(S_n(X))$. Since $KI(f^q, \eta_{n,m} z^q) = KI(\overset{*}{\eta}_{n,m} f^q, z^q)$ it follows that $f^q \in \Lambda_q(S_m(X), G)$ if and only if $KI(\overset{*}{\eta}_{n,m} f^q, z^q) = 0$ for all $z^q \in \Sigma^q(S_n(X))$, i.e. if $\overset{*}{\eta}_{n,m} f^q \in \Lambda_q(S_n(X), G)$.

35. Relations with products

We shall now show how the spherical cycles and their annihilators are related to the products.

THEOREM 35.1. *Let G_1 and G_2 be paired to G. Given a cohomology class $f^q \in H_q(X, G_1)$, $q > 0$, and a spherical homology class $z^{p+q} \in \Sigma^{p+q}(X, G_2)$, $p > 0$, we have*

$$f^q \cap z^{p+q} = 0.$$

PROOF. By definition we have $z^{p+q} = \sum g_i z_i^{p+q}$ where $z_i^{p+q} \in \Sigma^{p+q}(X)$. It is therefore sufficient to show that $f^q \cup z_i^{p+q} = 0$. Since $z_i^{p+q} \in \Sigma^{p+q}(X)$ there is a continuous mapping

$$\varphi: S^{p+q} \to X$$

and a homology class $c_i^{p+p} \in H^{p+q}(S^{p+q})$ such that

$$\varphi(c_i^{p+q}) = z_i^{p+q}.$$

Since the chain transformation

$$\varphi: S(S^{p+q}) \to S(X)$$

preserves the product it follows by (27.2) that

$$\varphi(\varphi^* f^q \cup c_i^{p+q}) = f^q \cup \varphi(c_i^{p+q}) = f^q \cup z_i^{p+q}.$$

Hence $f^q \cup z_i^{p+q}$ is the image under φ of a homology class in $H^p(S^{p+q})$. Since $0 < p < p + q$ we have $H^p(S^{p+q}) = 0$ and therefore $f^q \cup z_i^{p+q} = 0$.

THEOREM 35.2. *Let G_1 and G_2 be paired to G. Given two cohomology classes $f_1^p \in H_p(X, G_1)$ and $f_2^q \in H_q(X, G_2)$, $p > 0$, $q > 0$, we have*

$$f_1^p \cup f_2^q \in \Lambda_{p+q}(X, G).$$

PROOF. Let $z^{p+q} \in \Sigma^{p+q}(X)$. By ($\cap$ 2) we have

$$(f_1^p \cup f_2^q) \cap z^{p+q} = f_1^p \cap (f_2^q \cap z^{p+q})$$

but $f_2^q \cap z^{p+q} = 0$ by the previous theorem. Hence $(f_1^p \cup f_2^q) \cap z^{p+q} = 0$ and by ($\cap$ 4) $KI(f_1^p \cup f_2^q, z^{p+q}) = 0$. Since this holds for every $z^{p+q} \in \Sigma^{p+q}(X)$, Theorem 35.2 follows.

REMARK. Both Theorems 35.1 and 35.2 could be restated for any one of the complexes $S_n(X)$ where $n \leqq p + q$.

APPENDIX

CHAIN EQUIVALENCES AND NATURALITY

In the course of this paper we have encountered several chain equivalences

$$\tau: K_1 \to K_2 .$$

To be more specific we had

$$\alpha: K(P) \to k(P)$$

$$\beta: K(P) \to S(P)$$

$$\eta_{n+1,n}: S_{n+1}(X) \to S_n(X).$$

The latter was a chain equivalence under the assumption that $\pi_n(X) = 0$. For each of the transformations α, β, $\eta_{n+1,n}$ we have constructed a transformation $\bar{\alpha}$ or $\bar{\beta}$ or $\bar{\eta}_{n+1,n}$ such that $(\alpha, \bar{\alpha})$, $(\beta, \bar{\beta})$ and $(\eta_{n+1,n}, \bar{\eta}_{n+1,n})$ were equivalence pairs. We wish to remark here that while the definitions of α, β and $\eta_{n+1,n}$ were natural[22] and in a certain sense unique, the definition of the "inverses" $\bar{\alpha}$, $\bar{\beta}$ and $\bar{\eta}_{n+1,n}$ involved many choices and were by no means natural. If we restrict our attention to homology and cohomology groups, then the isomorphisms induced by $\bar{\alpha}$, $\bar{\beta}$ and $\bar{\eta}_{n+1,n}$ are inverses of those induced by α, β and $\eta_{n+1,n}$. Consequently the non-naturality of the definitions of $\bar{\alpha}$, $\bar{\beta}$ and $\bar{\eta}_{n+1,n}$ disappears once we pass to the homology and cohomology groups.

[22] The concept of naturality used here in its intuitive sense could be made quite precise and mathematical using the concept of a functor. See S. Eilenberg and S. MacLane, *Natural isomorphisms in group theory*, Proc. Nat. Acad., U. S. A., 28 (1942), pp. 537–543; and also a forthcoming paper entitled "*General theory of natural equivalences*" by the same authors.

This creates a curious situation when we compare the complexes $k(P)$ and $S(P)$. We have the natural chain transformations

$$k(P) \xleftarrow[\alpha]{} K(P) \xrightarrow[\beta]{} S(P).$$

The chain transformations

$$\beta\bar{\alpha}: k(P) \to S(P)$$
$$\alpha\bar{\beta}: S(P) \to k(P)$$

will only be natural in the weak sense, i.e. if we restrict our attention to the homology and cohomology groups.

The same situation repeats itself when we compare $K(P)$ and $S_n(P)$. We then have

$$K(P) \xrightarrow[\beta]{} S(P) \xleftarrow[\eta_n]{} S_n(P).$$

Concerning the products we remark that in all the complexes considered in this paper, except in the complex $k(P)$, the products were defined in a natural fashion. The chain transformations β and $\eta_{n+1,n}$ both preserved the products while the chain transformation α did not.

UNIVERSITY OF MICHIGAN.

ANNALS OF MATHEMATICS
Vol. 45, No. 3, July, 1944

REMARK ON A PAPER BY R. H. FOX[1]

BY HANS SAMELSON

(Received December 18, 1943)

"... In your paper "On Homotopy Type and Deformation Retracts" Ann. of Math. 44 (1943), (p. 40–50) (cited in the sequel as F) statement (1.3) seems incorrect; the equations given in F do not define a continuous deformation. Counterexamples can be given: let A be the set $1 \leqq r \leqq 2$ (r, φ polar coordinates in the plane), and define a deformation ξ of A by the equation

$$\xi_t(r, \varphi) = (r, \varphi + (2 - r)\cdot 2\pi\cdot t), \qquad 0 \leqq t \leqq 1.$$

The set B of fixed points of ξ_1 consists of $r = 1$ and $r = 2$; it is clearly impossible to deform ξ_0 into ξ_1, keeping all points of B fixed during the deformation.

Now in the proof of Theorem (1.4) in F you use (1.3) to prove the necessity of condition (i); the proof therefore has to be changed. This can be done in the following way, using an idea from F. The statement to be proved reads:

Let A be an ANR-set, and let the closed subset B be a deformation retract of A; then there exists a deformation ξ of A into B such that the points of B are fixed under each of the mappings ξ_t, $0 \leqq t \leqq 1$.

PROOF. The hypothesis that B be a deformation retract of A means that there exists a retraction of A onto B which is homotopic to the identity mapping of A; call the retraction r and the homotopy ρ, so that $r(A) = B$, $r^2 = r$, $\rho_0 =$ identity and $\rho_1 = r$.

Consider the closed subset $C = A \times [0] + B \times [0, 1] + A \times [1]$ of $A \times [0, 1]$. Define a homotopy η of C in the following way:

$$\begin{array}{lll} \alpha) & \eta_u(a, 0) = a & 0 \leqq u \leqq 1, a \in A \\ \beta) & \eta_u(b, t) = \rho_{(1-u)t}(b) & 0 \leqq u \leqq 1, b \in B, 0 \leqq t \leqq 1, \\ \gamma) & \eta_u(a, 1) = \rho_{1-u}(r(a)) & 0 \leqq u \leqq 1, a \in A. \end{array}$$

This is well defined: points $(b, 0)$ are mapped by α) into b and by β) into $\rho_0(b)$ which is equal to b; points $(b, 1)$ are mapped by β) into $\rho_{1-u}(b)$, by γ) into $\rho_{1-u}(r(b))$, and $r(b) = b$ for all $b \in B$.

It is a continuous mapping: That is clear for points $(a, 0)$ and $(a, 1)$ with $a \notin B$, and for points (b, t) with $0 < t < 1$. For any sequence $u_n \to u$, $a_n \to b$, $t_n \to 0$ use α) for the subsequence defined by $t_n = 0$, and β) for the subsequence defined by $t_n \neq 0$, and apply $\rho_0(b) = b$. Similarly for sequences $u_n \to u$, $a_n \to b$, $t_n \to 1$.

Now η_0 has an extension $\bar{\eta}_0$ to $A \times [0, 1]$, namely $\bar{\eta}_0(a, t) = \rho_t(a)$ (notice in γ) for $u = 0$ that $\rho_1(r(a)) = r^2(a) = r(a) = \rho_1(a)$). Therefore according to the well known extension theorem, A being an ANR-set, η_1 also has an extension $\bar{\eta}_1$

[1] Extract from a letter to R. H. Fox published upon suggestion from Fox.

to $A \times [0, 1]$ (cf. F, p. 42). This extension defines the required deformation ξ of A by $\xi_t(a) = \bar{\eta}_1(a, t)$. We have

$$\xi_0(a) = \bar{\eta}_1(a, 0) = a \text{ from } \alpha),$$

$$\xi_t(b) = \bar{\eta}_1(b, t) = \rho_0(b) = b \text{ for } b \in B \text{ and } 0 \leqq t \leqq 1 \text{ from } \beta) \text{ and}$$

$$\xi_1(a) = \bar{\eta}_1(a, 1) = \rho_0(r(a)) = r(a) \text{ from } \gamma).$$

This concludes the proof. The homotopy η can be loosely described this way: the paths $\rho_t(b)$ are contracted in themselves towards their origin, and the images of points $(a, 1)$ are moved along the paths $\rho_t(r(a))$ in reversed direction."

Syracuse University

Annals of Mathematics
Vol. 45, No. 3, July, 1944

THE IMBEDDING PROBLEM FOR MODULAR LATTICES

By M. Hall and R. P. Dilworth

(Received November 11, 1943)

1. Introduction

It is trivially true that an arbitrary lattice may be imbedded in a complemented lattice. We need only adjoin a unit and null elements if they do not already exist and a single element which is a complement of each of the elements not the unit of null element. For distributive lattices, the imbedding problem is not so trivial, but is contained in the representation theorem which asserts that any distributive lattice is isomorphic with a ring of sets (Birkhoff (1), Mac Neille (1)). The corresponding problem of imbedding a modular lattice in a complemented modular lattice is an outstanding problem in lattice theory. We exhibit here an example of a modular lattice which cannot be imbedded in any complemented modular lattice. However we will be concerned primarily with the isometric problem for finite dimensional modular lattices; that is, the problem of imbedding a finite dimensional modular lattice in a complemented modular lattice of the same dimensionality.

2. Notation and definitions

We will use the notation and terminology of G. Birkhoff "Lattice Theory" except that proper inclusion will be denoted by $a \supset b$, reserving $a > b$ to indicate that a covers b. It will be recalled that a lattice L is modular if

M: $a \supseteq b$ implies $a \cap (b \cup c) = b \cup (a \cap c)$ and complemented if

C: For each a there is at least one complement a' such that $a \cup a' = u$, $a \cap a' = z$ where u and z are respectively unit and null elements of the lattice.

Any complemented modular lattice is also relatively complemented, that is whenever $a \supseteq x \supseteq b$ there exists a y such that $x \cup y = a$, $x \cap y = b$.

Considerable use will be made of the following fundamental theorem on complemented modular lattices (Birkhoff (1) pp. 60).

Theorem 2.1. *Any finite-dimensional complemented modular lattice is the direct product of projective geometrics.*

It is understood in the statement of the theorem that a Boolean algebra of order two is the projective geometry of the void space and a single point where we may consider the geometric postulates on lines as satisfied by default.

3. Subdirect product decompositions

As a first reduction of the isometric imbedding problem, the lattice L will be represented as a subdirect product of lattices having a simple structure with respect to projectivity.

Definition 3.1. A modular lattice is said to be *projective* if every two prime quotients are projective (Birkhoff (1) pp. 37).

The following theorem was proved by Birkhoff for lattices of finite dimension. The proof given here is new and applies to any atomic lattice.

THEOREM 3.1. *Let L be an atomic modular lattice. Then L is a subdirect product of projective lattices.*

PROOF. Atomicity will be used in the following weak form: $a \supset b$ implies x exists such that $a \supseteq x > b$. Now let $\mathfrak{p}$ be a given prime quotient in L and let $\mathfrak{P}$ denote the set of all prime quotients projective to $\mathfrak{p}$. In L we define $a \sim b$ ($\mathfrak{P}$) if and only if there exists a chain of elements $a = a_1, a_2, \cdots, a_n = b$ such that $a_i \cup a_{i+1}/a_i \cap a_{i+1}$ contains *no* quotient projective to $\mathfrak{p}$. The relation $a \sim b$ is clearly an equivalence relation. We show that it is also a congruence relation. Let $a \sim b$ ($\mathfrak{P}$) and consider the chain $a \cup c = a_1 \cup c, a_2 \cup c, \cdots, a_n \cup c = b \cup c$. Suppose $(a_i \cup c) \cup (a_{i+1} \cup c)/(a_i \cup c) \cap (a_{i+1} \cup c)$ contains a quotient x/y projective to $\mathfrak{p}$. Then $x > y$ and $y \supseteq (a_i \cup c) \cap (a_{i+1} \cup c) \supseteq (a_i \cap a_{i+1}) \cup c$. Thus $y \supseteq c$ and $y \supseteq a_i \cap a_{i+1}$. Clearly $a_i \cup a_{i+1} \cup c \supseteq x \cup (a_i \cup a_{i+1}) \supseteq y \cup (a_i \cup a_{i+1}) \supseteq a_i \cup a_{i+1} \cup c$. Hence $x \cup (a_i \cup a_{i+1}) = y \cup (a_i \cup a_{i+1})$ and it follows that $x \cap (a_i \cup a_{i+1}) > y \cap (a_i \cup a_{i+1})$. But then $a_i \cup a_{i+1} \supseteq x \cap (a_i \cup a_{i+1}) > y \cap (a_i \cup a_{i+1}) \supseteq a_i \cap a_{i+1}$ and $x \cap (a_i \cup a_{i+1})/y \cap (a_i \cup a_{i+1})$ is projective to x/y and hence is projective to $\mathfrak{p}$ which is contrary to assumption. Thus $(a_i \cup c) \cup (a_{i+1} \cup c)/(a_i \cup c) \cap (a_{i+1} \cup c)$ contains no quotient projective to $\mathfrak{p}$ and we have $a \cup c \sim b \cup c$ ($\mathfrak{P}$). In a similar manner $a \cap c \sim b \cap c$ ($\mathfrak{P}$).

If $x, y, \cdots$ are elements of L, let $\{x\}_{\mathfrak{P}}, \{y\}_{\mathfrak{P}}, \cdots$ represent the congruence classes determined by $x, y, \cdots$ respectively. Since $a \sim b$ ($\mathfrak{P}$) is a congruence relation, if we define $\{x\}_{\mathfrak{P}} \cup \{y\}_{\mathfrak{P}} = \{x \cup y\}_{\mathfrak{P}}$ and $\{x\}_{\mathfrak{P}} \cap \{y\}_{\mathfrak{P}} = \{x \cap y\}_{\mathfrak{P}}$ the congruence classes form a modular lattice $L_{\mathfrak{P}}$. Now let a/b be a prime quotient projective to $\mathfrak{p}$. We shall show that $a \nsim b$ ($\mathfrak{P}$). For let $a = a_1, a_2, \cdots, a_n = b$ be a chain of elements such that $a_i \cup a_{i+1}/a_i \cap a_{i+1}$ contains no quotient projective to $\mathfrak{p}$. Now $a \cup (a_1 \cap a_2) = b \cup (a_1 \cap a_2)$ since otherwise $a \cup (a_1 \cap a_2)/b \cup (a_1 \cap a_2)$ is a quotient in $a_1 \cup a_2/a_1 \cap a_2$ which is projective to $\mathfrak{p}$. Hence $a \cap a_1 \cap a_2 > b \cap a_1 \cap a_2$. Now suppose it has been shown that $a \cap a_1 \cap \cdots \cap a_i > b \cap a_1 \cap \cdots \cap a_i$. Then $(a \cap a_1 \cap \cdots \cap a_i) \cup (a_i \cap a_{i+1}) = (b \cap a_1 \cap \cdots \cap a_i) \cup (a_i \cap a_{i+1})$ since otherwise $(a \cap a_1 \cap \cdots \cap a_i) \cup (a_i \cap a_{i+1})/(b \cap a_1 \cap \cdots \cap a_i) \cup (a_i \cap a_{i+1})$ is a prime quotient in $a_i \cup a_{i+1}/a_i \cap a_{i+1}$ which is projective to a/b and hence is projective to $\mathfrak{p}$ which contradicts our assumption. Hence $a \cap a_1 \cap \cdots \cap a_i \cap a_{i+1} > b \cap a_1 \cap \cdots \cap a_i \cap a_{i+1}$. By induction we get $a \cap a_1 \cap \cdots \cap a_n > b \cap a_1 \cap \cdots \cap a_n$. But then $a = b \cup (a_1 \cap \cdots \cap a_n)$. Now $a_n = b$ and hence $a = b$ which contradicts $a > b$. Thus $a \nsim b$ ($\mathfrak{P}$). It clearly follows that $\{a\}_{\mathfrak{P}} > \{b\}_{\mathfrak{P}}$ in $L_{\mathfrak{P}}$.

Now let $\{a\}_{\mathfrak{P}} > \{b\}_{\mathfrak{P}}$ and $\{c\}_{\mathfrak{P}} > \{d\}_{\mathfrak{P}}$ in $L_{\mathfrak{P}}$. Since $a \nsim b$, $a \cup b/a \cap b$ contains a quotient x/y which is projective to $\mathfrak{p}$. Then $\{a\}_{\mathfrak{P}} = \{a \cup b\}_{\mathfrak{P}} \supseteq \{x\}_{\mathfrak{P}} > \{y\}_{\mathfrak{P}} \supseteq \{a \cap b\}_{\mathfrak{P}} = \{b\}_{\mathfrak{P}}$ and hence $\{a\}_{\mathfrak{P}} = \{x\}_{\mathfrak{P}}$, $\{b\}_{\mathfrak{P}} = \{y\}_{\mathfrak{P}}$. In a similar manner $\{c\}_{\mathfrak{P}} = \{v\}_{\mathfrak{P}}$, $\{d\}_{\mathfrak{P}} = \{w\}_{\mathfrak{P}}$ where $v > w$ and v/w is projective to $\mathfrak{p}$. Thus x/y is projective to v/w and hence $\{a\}_{\mathfrak{P}}/\{b\}_{\mathfrak{P}}$ is projective to $\{c\}_{\mathfrak{P}}/\{d\}_{\mathfrak{P}}$ in $L_{\mathfrak{P}}$. It follows that $L_{\mathfrak{P}}$ is a projective lattice.

Let us set up the correspondence $a \to (\cdots, \{a\}_{\mathfrak{P}}, \cdots)$ where $\mathfrak{P}$ runs over all

sets of projective quotients. Clearly $a \cup b \to (\cdots, \{a\}_{\mathfrak{P}} \cup \{b\}_{\mathfrak{P}}, \cdots)$ and $a \cap b \to (\cdots, \{a\}_{\mathfrak{P}} \cap \{b\}_{\mathfrak{P}}, \cdots)$. Hence the correspondence is a homomorphism. But if $a \neq b$, then $a \cup b \supset a \cap b$ and x exists such that $a \cup b \supseteq x > a \cap b$. Let $\mathfrak{P}$ be the set of quotients projective to $x/a \cap b$. Then $\{x\}_{\mathfrak{P}} \neq \{a \cap b\}_{\mathfrak{P}}$ and hence $\{a \cup b\}_{\mathfrak{P}} \neq \{a \cap b\}_{\mathfrak{P}}$ and clearly $\{a\}_{\mathfrak{P}} \neq \{b\}_{\mathfrak{P}}$. Thus $(\cdots, \{a\}_{\mathfrak{P}}, \cdots) \neq (\cdots, \{b\}_{\mathfrak{P}}, \cdots)$ and the correspondence is an isomorphism. If $a > b$, then $\{a\}_{\mathfrak{P}} > \{b\}_{\mathfrak{P}}$ where $\mathfrak{P}$ is the set of prime quotients projective to a/b, and $\{a\}_{\mathfrak{P}'} = \{b\}_{\mathfrak{P}'}$ where $\mathfrak{P}'$ is any other set of projective prime quotients. Hence $(\cdots, \{a\}_{\mathfrak{P}}, \cdots) > (\cdots, \{b\}_{\mathfrak{P}}, \cdots)$ in the direct product lattice. Thus we have shown that L is isomorphic with an isometric sublattice of the direct product of the projective lattices $L_{\mathfrak{P}}$.

Now if L is a lattice of finite dimensions, by theorem 3.1 it is an isometric sublattice of the direct product $L_1 \times \cdots \times L_n$ where L_i is a projective lattice. Hence if we can imbed L_i isometrically in a complemented modular lattice M_i, then L will be isometrically imbedded in the complemented modular lattice $M = M_1 \times \cdots \times M_n$. Thus the imbedding problem for arbitrary modular lattices of finite dimension is reduced to the consideration of projective lattices of finite dimension.

We remark in this connection that modular lattices of dimension three or less can always be imbedded in a complemented modular lattice (Dilworth (2)).

For projective lattices, the problem is further reduced by the following theorem:

THEOREM 3.2. *If a projective lattice of finite dimensions can be imbedded isometrically in a complemented modular lattice M, then M is a projective geometry.*

PROOF. Since M is a complemented modular lattice of finite dimensions, it is a direct product of projective geometries $M = P_1 \times \cdots \times P_n$. Now let $a > b$ in L. Then $a = (a_1, \cdots, a_n)$ and $b = (b_1, \cdots, b_n)$. Since $a > b$ and the imbedding is isometric we have $a_i > b_i$ for some i and $a_j = b_j$ for $j \neq i$. But since every prime quotient in L is projective to a/b and the correspondence $a \to a_i$ is a homomorphism we have $c > d$ implies $c_i > d_i$. Thus the correspondence $a \to a_i$ is an isomorphism and L is imbedded in P_i. But since the original imbedding is isometric, P_i is the only component which is not null and hence $M = P_i$. That is, M is a projective geometry.

4. Counter-examples

In the construction of the counter examples we shall need the following lemma:

LEMMA 4.1. *Let L_1 and L_2 be lattices with unit elements u_1, u_2 and null elements, z_1, z_2 respectively. Let the quotient lattice u_1/a_1 of L_1 be isomorphic to the quotient lattice a_2/z_2 of L_2. If isomorphic elements are identified, then the set sum L of L_1 and L_2 is a lattice which contains L_1 and L_2 as sublattices and is modular if and only if L_1 and L_2 are modular.*

PROOF. Let a and b be any two elements of L. If both a and b are in L_1 or in L_2, then $a \cup b$ is the union in L_1 or L_2 and $a \cap b$ is cross-cut in L_1 or L_2 respectively. If a is in L_1 and b is in L_2, then $a \cup b = (a \cup z_2) \cup b$ where the first

union is in L_1 and the second, in L_2. Similarly $a \cap b = a \cap (u_1 \cap b)$ where the first cross-cut is in L_1 and the second, in L_2. It is readily verified that L is a lattice under these definitions of union and cross-cut. Clearly L_1 and L_2 are sublattices of L. That L is modular if and only if L_1 and L_2 are modular follows from Lemma 4.2 of Dilworth (1).

Consider a lattice L_1 whose diagram is given by Figure 1, where the quotient lattice a/z is the lattice of a non-Desarguesian plane. Since a/z and u/e are modular Lemma 4.1 assures us that L_1 is modular. Since every prime quotient in L_1 is projective to a/e, L_1 is a projective lattice and by Theorem 3.2 if it can be imbedded isometrically in a complemented modular lattice, then the complemented modular lattice is a projective geometry. Here u/z is three dimensional in the customary terminology of projective geometry (four dimensional as a lattice). But every plane in a projective 3-space must be Desarguesian (Veblen and Young (1)) and hence it must be impossible to imbed L_1 isometrically in a

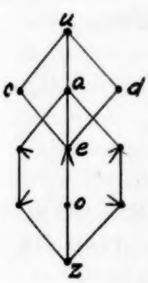

FIG. 1

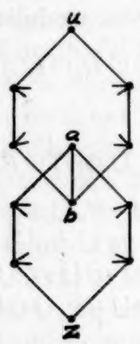

FIG. 2

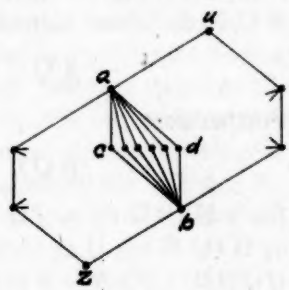

FIG. 3

complemented modular lattice. We have been able to show even more, namely that the following statement holds:

L_1 cannot be imbedded in any complemented modular lattice.

For since a/z is non-Desarguesian, there exists "points" O, A_1, B_1, C_1, A_2, B_2, C_2, A_3, B_3, C_3 and "lines" $e = OA_1A_2$, OB_1B_2, $f = OC_1C_2$, $A_1B_1C_3$, $A_2B_2C_3$, $A_1B_3C_1$, $b = A_2B_3C_2$, $A_3B_1C_1$, $A_3B_2C_2$ such that $A_1 \cup B_1 \cup C_1 = A_2 \cup B_2 \cup C_2 = A_3 \cup B_3 \cup C_3 = a$. Thus A_1, B_1, C_1 and A_2, B_2, C_2 are two triangles perspective from O, whose corresponding sides A_1B_1 and A_2B_2, etc. meet in three non-collinear points C_3, A_3, B_3.

Now suppose that u/z is a sublattice of a complemented modular lattice M. Let x be a relative complement of e in d/O and y be a relative complement of e in d/A_2 so that $x \cup e = y \cup d$, $x \cap e = O$, $y \cap e = A_2$. From these relations the following projectivities may be verified: f/z proj. a/A_2 proj. u/y proj. c/A_2 proj. u/b proj. x/z. Furthermore, under the projections $O \to e \to d \to e \to a \to O$. Let O_1 and O_2 be the image of C_1 and C_2 under the series of projections. Then

since $C_1 \cup O = C_2 \cup O = C_1 \cup C_2 = f$ and $C_1 \cap O = C_2 \cap O = C_1 \cap C_2 = Z$ we have $O_1 \cup O = O_2 \cup O = O_1 \cup O_2 = x$ and $O_1 \cap O = O_2 \cap O = O_1 \cap O_2 = z$. Let us set

$$A = (A_1 \cup O_1) \cap (A_2 \cup O_2)$$
$$B = (B_1 \cup O_1) \cap (B_2 \cup O_2)$$
$$C = (C_1 \cup O_1) \cap (C_2 \cup O_2).$$

Then $A \cup B = (A_1 \cup B_1 \cup O_1) \cap (A_2 \cup B_2 \cup O_2)$. For $e \cup A \cup B = e \cup A_1 \cup A_2 \cup A \cup B = e \cup [(A_1 \cup A_2 \cup O_1) \cap (A_1 \cup A_2 \cup O_2)] \cup B = [(e \cup O_1) \cap (e \cup O_2)] \cup B = d \cup B = d \cup O_1 \cup O_2 \cup [(B_1 \cup O_1) \cap (B_2 \cup O_2)] = d \cup [(B_1 \cup O_1 \cup O_2) \cap (B_2 \cup O_1 \cup O_2)] = d \cup [(B_1 \cup x) \cap (B_2 \cup x)] = d \cup [(B_1 \cup O \cup x) \cap (B_2 \cup O \cup x)] = d \cup B_1 \cup B_2 = d \cup a = u$ while $e \cap (A_1 \cup B_1 \cup O_1) \cap (A_2 \cup B_2 \cup O_2) = e \cap d \cap (A_1 \cup B_1 \cup O_1) \cap d \cap (A_2 \cup B_2 \cup O_2) = e \cap (A_1 \cup O_1) \cap (A_2 \cup O_2) = [A_1 \cup (e \cap O_1)] \cap [A_2 \cup (e \cap O_2)] = A_1 \cap A_2 = z$. Since $(A_1 \cup B_1 \cup O_1) \cap (A_2 \cup B_2 \cup O_2) \supseteq A \cup B$ the above formula follows from modularity. By symmetry

$$A \cup C = (A_1 \cup C_1 \cup O_1) \cap (A_2 \cup C_2 \cup O_2).$$

Furthermore

$$B \cup C = (B_1 \cup C_1 \cup O_1) \cap (B_2 \cup C_2 \cup O_2).$$

For $f \cup (B \cup C) = f \cup (C_1 \cup C_2) \cup [(C_1 \cup O_1) \cap (C_2 \cup O_2)] \cup B = f \cup [(C_1 \cup C_2 \cup O_1) \cap (C_1 \cup C_2 \cup O_2)] \cup B = [(f \cup O_1) \cap (f \cup O_2)] \cup B = [(f \cup O \cup O_1) \cap (f \cup O \cup O_2)] \cup B = f \cup x \cup B = f \cup O_1 \cup O_2 \cup [(B_1 \cup O_1) \cap (B_2 \cup O_2)] = f \cup [(B_1 \cup O_1 \cup O_2) \cap (B_2 \cup O_1 \cup O_2)] = f \cup [(B_1 \cup O \cup x) \cap (B_2 \cup O \cup x)] = f \cup B_1 \cup B_2 \cup x = a \cup x = a \cup e \cup x = a \cup d = u$ while $f \cap (B_1 \cup C_1 \cup O_1) \cap (B_2 \cup C_2 \cup O_2) = f \cap a \cap (B_1 \cup C_1 \cup O_1) \cap (B_2 \cup C_2 \cup O_2) = f \cap (B_1 \cup C_1) \cap (B_2 \cup C_2) = C_1 \cap C_2 = z$. Since $(B_1 \cup C_1 \cup O_1) \cap (B_2 \cup C_2 \cup O_2) \supseteq B \cup C$, the equality must hold by the modular law.

But then $A \cup B \supseteq (A_1 \cup B_1) \cap (A_2 \cup B_2) \supseteq C_3$ and similarly $A \cup C \supseteq B_3$, $B \cup C \supseteq A_3$. Thus $A \cup B \cup C \supseteq A_3 \cup B_3 \cup C_3 = a$. Hence $A_1 = A_1 \cap a = A_1 \cap (A \cup B \cup C) = A_1 \cap (A_1 \cup B_1 \cup O_1) \cap (A \cup B \cup C) = A_1 \cap [A \cup B \cup ((A_1 \cup B_1 \cup O_1) \cap C)] = A_1 \cap [A \cup B \cup ((A_1 \cup B_1 \cup O_1) \cap (C_1 \cup O_1) \cap (C_2 \cup O_2))] = A_1 \cap [A \cup B \cup ((O_1 \cup [C_1 \cap (A_1 \cup B_1)] \cap ((C_2 \cup O_2))] = A_1 \cap [A \cup B \cup (O_1 \cap (C_2 \cup O_2))] = A_1 \cap [A \cup B \cup (O_1 \cap d \cap (C_2 \cup O_2))] = A_1 \cap [A \cup B \cup (O_1 \cap O_2)] = A_1 \cap [A \cup B] = A_1 \cap (A_2 \cup B_2 \cup O_2) = A_1 \cap a \cap (A_2 \cup B_2 \cup O_2) = A_1 \cap (A_2 \cup B_2) = z$. But A_1 is a "point" of a/z and hence is not equal to z. Thus we have a contradiction and u/z cannot be imbedded in any complemented modular lattice.

Let L_2 be a lattice whose diagram is given by Figure 2 where u/b is a Desarguesian plane whose coordinatizing skew-field F is of characteristic p and a/Z is a Desarguesian plane whose coordinatizing skew-field K is of characteristic q and let $q \neq p$. Then, as with L_1, if L_2 can be imbedded isometrically in a complemented modular lattice, this lattice must be a projective 4-space G. But the

coordinatizing skew-field of G cannot have subfields F and K of different characteristics and we are led to a contradiction. Hence L_2 cannot be imbedded isometrically in a complemented modular lattice.

The third counter-example is the lattice L_3 whose diagram is Figure 3. Here a/z and u/b are isomorphic Desarguesian planes whose field F contains more than three elements and is not of characteristic 2. We may, for example, let F be the finite field of five elements. In general if F contains n elements, then in the plane which it determines there are $n + 1$ points on every line and $n + 1$ lines through every point. Hence L_3 may be constructed from two isomorphic planes a_1/z and u/b_2 where since a_1/b_1 and a_2/b_2 contain the same number of elements the identifications $a_1 = a_2$, $b_1 = b_2$ and any mapping of the intermediate elements of a_1/b_1 and a_2/b_2 will satisfy the conditions of Lemma 4.1, yielding the modular lattice L_3. Now every prime quotient in L_3 is projective to a/c and hence L_3 is a projective lattice and by Theorem 3.2 if it can be imbedded in a complemented modular lattice, this lattice must be a projective geometry.

Now Lemma 4.1 assures us that an arbitrary mapping of a_1/b_1 on a_2/b_2 will make L_3 modular, but we shall show that if L_3 is imbedded in a projective 3-space then this mapping of a_1/b_1 onto a_2/b_2 cannot be arbitrary, and that therefore the mappings of a_1/b_1 onto a_2/b_2 which are not permissible in this way yield lattices L_3 which cannot be isometrically imbedded in complemented modular lattices. This will depend on the construction of a harmonic line conjugate. If A_1, B_1, C_1 are three lines of the quotient a_1/b_1 then we may construct in a_1/z a line C_1^*, in a_1/b_1 which is the harmonic conjugate of C_1 with respect to A_1 and B_1. Here $C_1^* \neq A_1$, B_1 always and $C_1^* \neq C_1$ if the characteristic of F is not 2. Similarly if A_2, B_2, C_2 are three lines of a_2/b_2, we may construct in u/b_2 a line C_2^* in a_2/b_2 which is the harmonic conjugate of C_2 with respect to A_2 and B_2. Here if u/z is a projective 3-space we must have $C_1^* = C_2^*$ and hence if a_1/b_1 is mapped onto a_2/b_2 so that $A_1 \rightleftarrows A_2$, $B_1 \rightleftarrows B_2$, $C_1 \rightleftarrows C_2$, then we must also map C_1^* onto C_2^* if u/z can be imbedded in a 3-space. But if a/b contains more than four elements and F is not of characteristic 2, this excludes certain mappings of a_1/b_1 onto a_2/b_2.

THEOREM 4.1. *Given a projective 3-space S and a plane π containing a point P, let A, B, C, be any three lines through P lying in π. Let R_1 and S_1 be two lines in π such that R_1, S_1, C are concurrent in a point Q_1 different from P. Construct $M_1 = (A \cap R_1) \cup (B \cap S_1)$, $N_1 = (A \cap S_1) \cup (B \cap R_1)$ and $C_1^* = P \cup (M_1 \cap N_1)$. Let R_2 and S_2 be two lines through P such that R_2, S_2, and C lie in a plane Q_2 different from π. Construct $M_2 = (A \cup R_2) \cap (B \cup S_2)$, $N_2 = (A \cup S_2) \cap (B \cup R_2)$ and $C_2^* = \pi \cap (M_2 \cup N_2)$. Then C_1^*, the harmonic conjugate from below of C with respect to A and B, is independent of the choice of R_1 and S_1; C_2^*, the harmonic conjugate from above of C with respect to A and B is independent of the choice of R_2 and S_2; and $C_1^* = C_2^*$.*

PROOF. We may introduce coordinates in S from the appropriate skew-field F where points are given by homogeneous coordinates (x_1, x_2, x_3, x_4) and planes have equations $u_1x_1 + u_2x_2 + u_3x_3 + u_4x_4 = 0$. By an appropriate choice

of the frame of reference we may take P as $(1, 0, 0, 0)$, π as $x_4 = 0$, and A as $x_2 = 0$, $x_4 = 0$, B as $x_3 = 0$, $x_4 = 0$, and C as $x_2 + x_3 = 0$, $x_4 = 0$.

Here we may take R_1 as $x_1 + \alpha x_2 + \beta x_3 = 0, x_4 = 0$ and S_1 as $x_1 + (\alpha + \gamma)x_2 + (\beta + \gamma)x_3 = 0$, $x_4 = 0$. Then $A \cap R_1 = (-\beta, 0, 1, 0)$, $B \cap S_1 = (-\alpha - \gamma, 1, 0, 0)$. Hence $M_1 = (A \cap R_1) \cup (B \cap S_1)$ is $x_1 + (\alpha + \gamma)x_2 + \beta x_3 = 0$, $x_4 = 0$. Then $B \cap R_1$ is $(-\alpha, 1, 0, 0)$, $A \cap S_1$ is $(-\beta - \gamma, 0, 1, 0)$ and $N_1 = (A \cap S_1) \cup (B \cap R_1)$ is $x_1 + \alpha x_2 + (\beta + \gamma)x_3 = 0$, $x_4 = 0$. Then $M_1 \cap N_1$ is $(-\alpha - \beta - \gamma, 1, 1, 0)$ and $C_1^* = P \cup (M_1 \cap N_1)$ is $x_2 - x_3 = 0$, $x_4 = 0$. Hence C_1^* is independent of the choice of R_1 and S_1.

We may take

$$R_2 \begin{cases} x_2 + x_3 + \alpha x_4 = 0 \\ \qquad x_3 + \beta x_4 = 0 \end{cases} \qquad S_2 \begin{cases} x_2 + x_3 + \alpha x_4 = 0 \\ \qquad x_3 + \gamma x_4 = 0 \end{cases}$$

since S is a line coplanar with R_2 and C. Then we have

$$\left.\begin{array}{ll} A \cup R_2 & x_2 + (\alpha - \beta)x_4 = 0 \\ B \cup S_2 & x_3 + \gamma x_4 \qquad\quad = 0 \end{array}\right\} M_2 = (A \cup R_2) \cap (B \cup S_2)$$

$$\left.\begin{array}{ll} A \cup S_2 & x_2 + (\alpha - \gamma)x_4 = 0 \\ B \cup R_2 & x_3 + \beta x_4 \qquad\quad = 0 \end{array}\right\} N_2 = (A \cup S_2) \cap (B \cup R_2)$$

Hence $M_2 \cup N_2$ is $x_2 - x_3 + (\alpha - \beta - \gamma)x_4 = 0$ and $C_2^* = (M_2 \cup N_2) \cap \pi$ is $x_2 - x_3 = 0$, $x_4 = 0$. Finally C_2^* is independent of the choice of R_2 and S_2 and $C_2^* = C_1^* = C^*$. Note that always $C^* \neq A, B$ and that $C^* \neq C$ if the characteristic of F is different from 2.

YALE UNIVERSITY.

REFERENCES

1. G. BIRKHOFF, *Lattice Theory*, Amer. Math. Soc. Coll. Pub., vol. 25.
2. R. P. DILWORTH, *The arithmetical theory of Birkhoff lattices*, Duke Math. Journal vol. 8 (1941) pp. 286–299.
3. ——— *On dependence relations in a semi-modular lattice*, submitted to the Duke Math. Journal.
4. H. M. MACNEILLE, *Partially ordered sets*, Trans. Amer. Math. Soc. vol. 42 (1937) pp. 416–460.
5. VEBLEN AND YOUNG, *Projective Geometry*, vol. I.

ANNALS OF MATHEMATICS
Vol. 45, No. 3, July, 1944

ON THE CONVERGENCE OF TRIGONOMETRICAL INTERPOLATION AT EQUI-DISTANT KNOTS

BY I. P. NATANSON

(Received October 12, 1943)

§1. Introduction

A trigonometrical polynomial of order n is a function of the form

$$T_n(x) = A + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx). \tag{1}$$

If $T_n(x)$ coincides at the points

$$x_0, x_1, x_2, \cdots, x_{2n} \qquad (0 \leqq x_k < 2\pi) \tag{2}$$

with some function $f(x)$, then $T_n(x)$ is called the interpolation polynomial of the function $f(x)$ at the knots (2).

In particular, the points

$$x_k^{(n)} = \frac{2\pi}{2n+1} k \qquad (k = 0, 1, \cdots, 2n) \tag{3}$$

are called equi-distant knots.

It is easily seen that the interpolation polynomial of these knots has the form

$$T_n(x) = \frac{1}{2n+1} \sum_{k=0}^{2n} f(x_k^{(n)}) \frac{\sin \frac{2n+1}{2} (x_k^{(n)} - x)}{\sin \frac{x_k^{(n)} - x}{2}} \tag{4}$$

where $f(x)$ is the function interpolated.

Naturally the question arises, under what conditions the polynomial $T_n(x)$ converges to the function $f(x)$ when n is increasing.

Papers dealing with this problem generally make some assumptions concerning the nature of the function $f(x)$ in the whole segment $[0, 2\pi]$. Such a treatment makes it possible to apply the results established in the theory of the best approximation of functions and leads to different tests of **uniform** convergence of $T_n(x)$ towards $f(x)$. These tests prove very similar to those of uniform convergence of Fourier-series.

Yet there is also a second possibility of treatment which it is natural to call a *local* one. Namely, we may examine properties of $f(x)$ near a fixed point x_0 which will cause the convergence of $T_n(x_0)$ towards $f(x_0)$. I know of only one paper treating the problem in this way, namely the classic mémoir by Vallée-Poussin.[1]

Vallée-Poussin has proved that the equality

$$\lim_{n\to\infty} T_n(x_0) = f(x_0) \tag{5}$$

[1] Ch.—J. de la Vallée-Poussin. "Sur la convergence de formules d'interpolation etc." Bull. de l'Acad. de Belgique, 1908.

holds whenever $f(x)$, being continuous at the point x_0 itself, is of bounded variation near this point. Consequently, the well known Dirichlet-Jordan test for the convergence of Fourier-series is at the same time a test of convergence of the interpolation process. Furthermore, Vallée-Poussin has established the equality (5) in the case where a finite $f'(x_0)$ exists or, more generally, in the case where all derivative numbers of $f(x)$ are finite at the point x_0. Thus the analogy between the theory of Fourier-series and the theory of interpolation at the knots (3) remains valid for the local treatment of the problem as well.

One of well known local tests of convergence is the Dini test requiring convergence of the integral

$$\int_0^{2\pi} \left| \frac{f(x) - f(x_0)}{x - x_0} \right| dx. \tag{6}$$

What has been said before naturally suggests the problem of extending this test to the theory of interpolation. This problem forms the main object of the present paper which, consequently, may be understood as an attempt to continue and develop the ideas of Vallée-Poussin.

The main result of this paper, as proved in §3, consists in establishing a theorem running as follows:

THEOREM. *Let $f(x)$ be a function integrable (R) in $[0, 2\pi]$, and let $0 < x_0 < 2\pi$. If there exists a function $\varphi(x)$ increasing in $[0, x_0]$, decreasing in $[x_0, 2\pi]$, and satisfying the conditions*

$$\left| \frac{f(x) - f(x_0)}{x - x_0} \right| \leq \varphi(x)$$

$$\int_0^{2\pi} \varphi(x)\, dx < +\infty,$$

then equality (5) is true.

The assumption of this theorem is heavier than the condition of the Dini-test. It is interesting to decide whether the Dini-test itself

$$\int_0^{2\pi} \left| \frac{f(x) - f(x_0)}{x - x_0} \right| dx < \infty \tag{7}$$

is sufficient for the validity of equality (5).

In §4, I show that test (7) is not sufficient for (5), even if only continuous functions are considered. Finally, in §5, I analyse some analogue to the Abel-Poisson process of summing a Fourier-series. Introductory lemmas that will be needed in the sequel are gathered in §2.

§2. Lemmas

LEMMA 1. *Put* $x_k^{(n)} = \dfrac{2\pi}{2n+1} k$ $(k = 0, \pm 1, \pm 2, \cdots)$ *and let p and m be integers* $(p \leq m)$. *Set, further,*

$$C_{n,p,m} = \sum_{k=p}^{m} \cos nx_k^{(n)}, \qquad S_{n,p,m} = \sum_{k=p}^{m} \sin nx_k^{(n)}.$$

Then

$$|C_{n,p,m}| < \frac{2}{\sqrt{3}}, \qquad |S_{n,p,m}| < \frac{2}{\sqrt{3}}. \tag{8}$$

PROOF. If we consider the sum

$$E_{n,p,m} = \sum_{k=p}^{m} e^{inx_k^{(n)}}$$

then $C_{n,p,m}$ and $S_{n,p,m}$ will be the real and imaginary parts of this sum. But as the numbers $x_k^{(n)}$ form an arithmetical progression, it is easily seen that

$$E_{n,p,m} = \frac{e^{inx_p^{(n)}} - e^{inx_{m+1}^{(n)}}}{1 - e^{inx_1^{(n)}}}$$

whence

$$|E_{n,p,m}| < \frac{2}{|1 - e^{i(2n\pi)/(2n+1)}|} = \frac{1}{\sin \dfrac{n\pi}{2n+1}}.$$

The expression

$$\sin \frac{n\pi}{2n+1}$$

attains its least value at $n = 1$, this value being $\sqrt{3}/2$. Thus the lemma is proved.

LEMMA 2. *Let a bounded function $f(x)$ be given in an arbitrary segment $[a, b]$, and let m and ω denote respectively the precise upper bound of the modulus of this function and its oscillation in $[a, b]$.*

On putting

$$C_n(f) = \frac{1}{2n+1} \sum_{k \in L_n} f(x_k^{(n)}) \cos nx_k^{(n)}, \qquad S_n(f) = \frac{1}{2n+1} \sum_{k \in L_n} f(x_k^{(n)}) \sin nx_k^{(n)}$$

where L_n is the set of all values of k for which $x_k^{(n)}$ falls into $[a, b]$, we obtain the inequalities:

$$|C_n(f)| \leqq \omega(b-a) + \frac{\omega + 2m}{2n+1}, \qquad |S_n(f)| \leqq \omega(b-a) + \frac{\omega + 2m}{2n+1}. \tag{9}$$

PROOF. Let c be an arbitrary point of $[a, b]$. Then

$$C_n(f) = \frac{1}{2n+1} \sum_{L_n} \{f(x_k^{(n)}) - f(c)\} \cos nx_k^{(n)} + \frac{f(c)}{2n+1} \sum_{L_n} \cos nx_k^{(n)}.$$

Hence, using (8), we get:

$$|C_n(f)| \leqq \frac{\omega \tau_n}{2n+1} + \frac{m}{2n+1} \cdot \frac{2}{\sqrt{3}} \tag{10}$$

where τ_n denotes the number of k − values belonging to L_n.

But, as the distance between two neighboring knots $x_k^{(n)}$ is $(2\pi)/(2n+1)$, the distance between the border knots belonging to the segment $[a, b]$ is

$$((2\pi)/(2n+1))(\tau_n - 1),$$

whence

$$\frac{\tau_n}{2n+1} \leqq \frac{b-a}{2\pi} + \frac{1}{2n+1}.$$

Combining this with (10) we get

$$| C_n(f) | \leqq \frac{\omega(b-a)}{2\pi} + \frac{\omega + m\frac{2}{\sqrt{3}}}{2n+1},$$

from which (9) follows at once. Evidently the estimate (9) remains valid also with respect to the semi-segment $[a, b)$ open on the right.

LEMMA 3. *Let the function $f(x)$ be integrable (R) in the segment $[a, b]$. Then, with the notation of lemma 2,*

$$\lim_{n\to\infty} C_n(f) = \lim_{n\to\infty} S_n(f) = 0.$$

PROOF. Take an arbitrary $\epsilon > 0$ and decompose the segment $[a, b]$ by the points

$$Z_0 = a < Z_1 < Z_2 < \cdots < Z_s = b$$

in parts $[Z_k, Z_{k+1}]$ so small that

$$\sum_{k=0}^{s-1} \omega_k(Z_{k+1} - Z_k) < \epsilon,$$

where ω_k is the oscillation of $f(x)$ in $[Z_k, Z_{k+1}]$. Applying estimate (9) to each of the intervals

$$[Z_0, Z_1), [Z_1, Z_2), \cdots, [Z_{s-2}, Z_{s-1}), [Z_{s-1}, Z_s]$$

we obtain

$$| C_n(f) | \leqq \sum_{k=0}^{s-1} \omega_k(Z_{k+1} - Z_k) + \frac{\sum_{k=0}^{s-1} (\omega_k + 2m_k)}{2n+1}$$

where m_k is the upper bound of $| f(x) |$ in $[Z_k, Z_{k+1}]$. Hence

$$| C_n(f) | < \epsilon + \frac{A(\epsilon)}{2n+1}$$

where $A(\epsilon) = \sum_{k=0}^{s-1} (\omega_k + 2m_k)$. Taking n sufficiently large we have

$$| C_n(f) | < 2\epsilon.$$

Thus the lemma is proved.

LEMMA 4. *Let $f(x)$ be a function integrable (R) in the segment $[0, 2\pi]$ which contains the point $x_0 (0 < x_0 < 2\pi)$, and let $[a, b]$ be a subsegment of $[0, 2\pi]$ that does not contain x_0. Furthermore, let L_n denote the set of all indices k for which $a \leqq x_k^{(n)} \leqq b$.*

On putting

$$U_n(f) = \frac{1}{2n+1} \sum_{L_n} f(x_k^{(n)}) \frac{\sin \frac{2n+1}{2} (x_k^{(n)} - x_0)}{\sin \frac{x_k^{(n)} - x_0}{2}}$$

we have

$$\lim_{n \to \infty} U_n(f) = 0.$$

PROOF. The function $\sin \frac{1}{2}(x - x_0)$ does not become zero in the segment $[a, b]$, and therefore the function

$$F(x) = \frac{f(x)}{\sin \frac{x - x_0}{2}}$$

is integrable (R). But

$$U_n(f) = \frac{1}{2n+1} \sum_{L_n} F(x_k^{(n)}) \sin (n + \tfrac{1}{2})(x_k^{(n)} - x_0),$$

hence we obtain

$$U_n(f) = \frac{1}{2n+1} \sum_{L_n} F(x_k^{(n)}) \Big[\sin n(x_k^{(n)} - x_0) \cos \frac{x_k^{(n)} - x_0}{2} + \cos n(x_k^{(n)} - x_0) \sin \frac{x_k^{(n)} - x_0}{2} \Big].$$

Writing[2]

$$F_1(x) = F(x) \cos \frac{x - x_0}{2}, \qquad F_2(x) = F(x) \sin \frac{x - x_0}{2},$$

we obtain

$$U_n(f) = \frac{1}{2n+1} \sum_{L_n} F_1(x_k^{(n)}) \sin n(x_k^{(n)} - x_0) + \frac{1}{2n+1} \sum_{L_n} F_2(x_k^{(n)}) \cos n(x_k^{(n)} - x_0).$$

[2] Evidently $F_2(x) = f(x)$, but this is not important.

The two sums obtained are quite similar to each other. Taking for instance the first of them and using the notation of Lemma 2 we have

$$\frac{1}{2n+1}\sum_{L_n} F_1(x_k^{(n)}) \sin n(x_k^{(n)} - x_0) = S_n(F_1)\cos nx_0 - C_n(F_1)\sin nx_0,$$

and thus Lemma 4 is reduced to Lemma 3.

REMARK. From Lemma 4 it follows that, if functions $f(x)$ and $g(x)$ coincide in an arbitrarily small neighborhood of a point x_0, their interpolation polynomials $T_n[f; x_0]$ and $T_n[g; x_0]$ at the point x_0 itself satisfy the condition

$$\lim_{n\to\infty} \{T_n[f; x_0] - T_n[g; x_0]\} = 0.$$

Hence, Riemann's localisation principle remains valid in the case of interpolation at knots (3). It is interesting to combine this fact, (stated for the first time by Vallée-Poussin) with S. N. Bernstein's theorem concerning the divergence of Newton's interpolation formula for the function $|x|$ at all points of the segment $[-1, +1]$. The result of S. N. Bernstein shows that in the case of parabolical interpolation at equi-distant knots there is no principle of localisation, the function $|x|$ coinciding in $[0, 1]$ with the function x for which the process is convergent. It might be interesting to derive general characteristics of those interpolation processes for which the principle of localisation is true.

LEMMA 5. *Let a positive increasing function $\varphi(x)$ be given in the segment $[a, b]$. If*

$$\int_a^b \varphi(x)\,dx < +\infty$$

then

$$\lim_{x\to b} \varphi(x)(b - x) = 0.$$

PROOF. The lemma follows from the obvious relations

$$\varphi(x)(b - x) \leqq \int_x^b \varphi(t)\,dt,$$

$$\lim_{x\to b}\int_x^b \varphi(t)\,dt = 0.$$

§3. The main theorem and its consequences

THEOREM 1. *Let the function $f(x)$ be given and integrable (R) in the segment $[0, 2\pi]$, and let*

$$T_n(x) = \frac{1}{2n+1}\sum_{k=0}^{2n} f(x_k^{(n)}) \frac{\sin\frac{2n+1}{2}(x_k^{(n)} - x)}{\sin\frac{x_k^{(n)} - x}{2}}$$

be its interpolation polynomial at knots (3).

If $0 < x_0 < 2\pi$, *and if there exists a function* $\varphi(x)$ *increasing in the segment* $[0, x_0]$, *decreasing in the segment* $[x_0, 2\pi]$, *and satisfying the conditions*

$$\left| \frac{f(x) - f(x_0)}{x - x_0} \right| < \varphi(x) \tag{11}$$

$$\int_0^{2\pi} \varphi(x)\, dx < +\infty \tag{12}$$

then

$$\lim_{n \to \infty} T_n(x_0) = f(x_0). \tag{13}$$

PROOF. From inequality (11) and Lemma (5) it follows that at the point x_0 the function $f(x)$ is continuous.

This stated, we choose an arbitrary $\epsilon > 0$ and take $\delta > 0$ so small that

$$\int_{x_0 - \delta}^{x_0 + \delta} \varphi(x)\, dx < \epsilon \tag{14}$$

and that the inequality $| x - x_0 | < \delta$ implies the inequality

$$| f(x) - f(x_0) | < \epsilon. \tag{15}$$

By virtue of the obvious equality

$$1 = \frac{1}{2n + 1} \sum_{k=0}^{2n} \frac{\sin \frac{2n+1}{2} (x_k^{(n)} - x_0)}{\sin \frac{x_k^{(n)} - x_0}{2}}$$

we obtain

$$T_n(x_0) - f(x_0) = \frac{1}{2n + 1} \sum_{k=0}^{2n} \{f(x_k^{(n)}) - f(x_0)\} \frac{\sin \frac{2n+1}{2} (x_k^{(n)} - x_0)}{\sin \frac{x_k^{(n)} - x_0}{2}}.$$

We now assume that

$$x_m^{(n)} < x_0 < x_{m+1}^{(n)}.$$

(If, for some n, the point x_0 coincides with one of the knots, the $T_n(x_0) = f(x_0)$, and such n may be excluded from consideration). We examine only the sum

$$r_n = \frac{1}{2n + 1} \sum_{k=0}^{m} \{f(x_k^{(n)}) - f(x_0)\} \frac{\sin \frac{2n+1}{2} (x_k^{(n)} - x_0)}{\sin \frac{x_k^{(n)} - x_0}{2}},$$

as the sum $\sum_{m+1}^{2n}$ is quite analogous.

If
$$x_{p-1}^{(n)} \leqq x_0 - \delta < x_p^{(n)}$$
and
$$r_n' = \frac{1}{2n+1} \sum_{k=0}^{p-1} \{f(x_k^{(n)}) - f(x_0)\} \frac{\sin \frac{2n+1}{2} (x_k^{(n)} - x_0)}{\sin \frac{x_k^{(n)} - x_0}{2}}$$
then, by Lemma 4,
$$\lim_{n\to\infty} r_n' = 0$$
and for $n > n_0$ we shall have

(16)
$$| r_n' | < \epsilon.$$
It remains to estimate the sum
$$r_n'' = \frac{1}{2n+1} \sum_{k=p}^{m} \{f(x_k^{(n)}) - f(x_0)\} \frac{\sin \frac{2n+1}{2} (x_k^{(n)} - x_0)}{\sin \frac{x_k^{(n)} - x_0}{2}}.$$

Its modulus is not greater than
$$\frac{1}{2n+1} \sum_{k=p}^{m-1} \left| \frac{f(x_k^{(n)}) - f(x_0)}{\sin \frac{x_k^{(n)} - x_0}{2}} \right| + \frac{1}{2n+1} \, |f(x_m^{(n)}) - f(x_0)|$$
$$\cdot \left| \frac{\sin \frac{2n+1}{2} (x_m^{(n)} - x_0)}{\sin \frac{x_m^{(n)} - x_0}{2}} \right|.$$

Observing that
$$| \sin n\alpha | \leqq n \, | \sin \alpha |$$
$$\sin \alpha > \frac{2}{\pi} \alpha \qquad \left(0 < \alpha < \frac{\pi}{2}\right)$$
and taking into account relation (15), we see that
$$| r_n'' | < \frac{\pi}{2n+1} \sum_{k=p}^{m-1} \left| \frac{f(x_k^{(n)}) - f(x_0)}{x_k^{(n)} \; x_0} \right| + \epsilon.$$
This and (11) imply

(17)
$$| r_n'' | < \frac{\pi}{2n+1} \sum_{k=p}^{m-1} \varphi(x_k^{(n)}) + \epsilon.$$

But, as the function $\varphi(x)$ is increasing, we have

$$\frac{2\pi}{2n+1}\varphi(x_k^{(n)}) < \int_{x_k^{(n)}}^{x_{k+1}^{(n)}} \varphi(x)\,dx. \qquad (k = p, p+1, \cdots, m-1)$$

From this it follows that

$$|r_n''| < \tfrac{3}{2}\epsilon.$$

Hence, using (16), we get for $n > n_0$,

$$|r_n| < \tfrac{5}{2}\epsilon.$$

Thus the theorem is proved.

THEOREM 2. *If at the point x_0 there exists a finite derivative $f'(x_0)$ or if*

$$|f(x) - f(x_0)| \leqq K|x - x_0|$$

then equality (13) *is true.*

In fact, the constant K may be taken for $\varphi(x)$. As stated in the introduction, Theorem 2 was established by Vallée-Poussin.

THEOREM 3. *If*

$$|f(x) - f(x_0)| \leqq K|x - x_0|^{\alpha} \qquad (\alpha > 0)$$

then (13) *is true.*

In fact, here we set $\varphi(x) = K|x - x_0|^{\alpha-1}$.

THEOREM 4. *If*

$$|f(x) - f(x_0)| \leqq \frac{K}{|\ln|x - x_0||^{1+\alpha}} \qquad (\alpha > 0)$$

then (13) *is true.*

In fact, here

$$\varphi(x) = \frac{K}{|x - x_0|\cdot|\ln|x - x_0||^{1+\alpha}}.$$

THEOREM 5.[3] *If*

$$|f(x) - f(x_0)|$$

$$\leqq \frac{K}{\ln\frac{1}{|x-x_0|}\cdot\ln_2\frac{1}{|x-x_0|}\cdots\ln_m\frac{1}{|x-x_0|}\cdot\left(\ln_{m+1}\frac{1}{|x-x_0|}\right)^{1+\alpha}} \qquad (\alpha > 0)$$

then (13) *is true.*

[3] As usual,

$$\ln_1 x = \ln x, \qquad \ln_{m+1} x = \ln(\ln_m x).$$

In fact, here we put

$$\varphi(x) = \frac{k}{|x - x_0| \cdot |\ln|x - x_0|| \cdot |\ln_2|x - x_0|| \cdots |\ln_m|x - x_0|| \cdot |\ln_{m+1}|x - x_0||^{1+\epsilon}}.$$

In all these theorems the function $f(x)$ is supposed to be integrable (R) in the segment $[0, 2\pi]$ and the point x_0 to be an internal point of this segment, $0 < x_0 < 2\pi$. The extension of these theorems to the case $x_0 = 0$ or $x_0 = 2\pi$ is quite trivial.

§4. An example of failure of the Dini test

The conditions for the convergence of the interpolation process established by Theorems 2, 3, 4, 5 are at the same time conditions for the convergence of the Fourier-series of the function $f(x)$ at x_0. Similarly, the Fourier-series for $f(x)$ converges to it in the case $x = x_0$, if the conditions of the main Theorem 1 are satisfied. But in the theory of Fourier-series it is not necessary to suppose that the ratio $\left|\frac{f(x) - f(x_0)}{x - x_0}\right|$ has an integrable monotonous majorant; it is sufficient to require the integrability of the ratio itself.

It is natural to investigate whether condition

$$\int_0^{2\pi} \left|\frac{f(x) - f(x_0)}{x - x_0}\right| dx < +\infty \tag{18}$$

is sufficient also for the convergence of the interpolation process. It is quite clear that for discontinuous functions conditions (18) is not sufficient; but the example we are going to analyze will prove its insufficiency for a case of a continuous function.

EXAMPLE. Let $n > 1$ be an integer. Let σ_n denote the greatest odd number $\leq \sqrt{n}$, and K_n the set of the odd numbers $1, 3, 5, \cdots, \sigma_n$.

We define a function $\varphi_n(x)$ by the conditions:

$$\varphi_n(x) = 0 \quad \text{for} \quad 0 \leq x \leq \pi$$

$$\varphi_n(x_{n+1}^{(n)}) = \varphi_n(x_{n+3}^{(n)}) = \cdots = \varphi_n(x_{n+\sigma_n}^{(n)}) = \lambda_n \tag{19}$$

where

$$\lambda_n = \frac{(2n+1)\sqrt{\ln \sigma_n}}{\sum_{i \in k_n} \frac{1}{\sin\left(\frac{2i-1}{2n+1} \cdot \frac{\pi}{2}\right)}} \tag{20}$$

We surround each of the points $x_{n+1}^{(n)}, x_{n+3}^{(n)}, \cdots, x_{n+\sigma_n}^{(n)}$ by an interval

$$(x_{n+i}^{(n)} - \delta_n, \quad x_{n+i}^{(n)} + \delta_n) \qquad (i = 1, 3, \cdots, \sigma_n) \tag{21}$$

so small that these intervals have no points in common and do not contain the knots $x_{n+2k}^{(n)}$, and that

$$x_{n+1}^{(n)} - \delta_n > \pi. \tag{22}$$

Evidently, these conditions will be satisfied if we choose

$$\delta_n < \frac{\pi}{2n+1} \tag{23}$$

but, in addition to (23), we yet require that

$$\int_{x_{n+i}^{(n)}-\delta_n}^{x_{n+1}^{(n)}+\delta_n} \frac{dx}{x-\pi} < \frac{1}{2^n \lambda_n \sigma_n} \qquad (i = 1, 3, \cdots, \sigma_n). \tag{24}$$

This attained, we make $\varphi_n(x)$ zero at the end points of the intervals (21), linear in the segments

$$[x_{n+i}^{(n)} - \delta_n, x_{n+i}^{(n)}] \quad \text{and} \quad [x_{n+i}^{(n)}, x_{n+i}^{(n)} + \delta_n], \qquad (i = 1, 3, \cdots \sigma_n)$$

and zero at all other points of $[0, 2\pi]$.

The function $\varphi_n(x)$ thus defined is non-negative, continuous, and by (24) we have

$$\int_{\pi}^{2\pi} \frac{\varphi_n(x)}{x-\pi}\, dx < \frac{1}{2^n}. \tag{25}$$

Furthermore, $\varphi_n(\pi) = 0$ and $\varphi_n'(\pi) = 0$, the latter derivative existing.

Denoting by $T_n[f; x]$ the interpolation polynomial for the function $f(x)$ at the knots $x_k^{(n)}$ we obtain

$$T_n[\varphi_n ; \pi] = \frac{1}{2n+1} \sum_{k=0}^{2n} \varphi(x_k^{(n)}) \frac{\sin \frac{2n+1}{2}(x_k^{(n)} - \pi)}{\sin \frac{x_k^{(n)} - \pi}{2}},$$

from which it follows that

$$T_n[\varphi_n ; \pi] = \frac{\lambda_n}{2n+1} \sum_{i \in k_n} \frac{\sin \frac{2n+1}{2}\left(\frac{2n+2i}{2n+1} - 1\right)\pi}{\sin\left(\frac{2i-1}{2n+1}\cdot\frac{\pi}{2}\right)}$$

and, by (20)

$$T_n[\varphi_n ; \pi] = \sqrt{\ln \sigma_n}. \tag{26}$$

Also, the inequality

$$0 < \sin \alpha < \alpha \qquad \left(0 < \alpha < \frac{\pi}{2}\right)$$

gives

$$\lambda_n < \frac{(2n+1)\sqrt{\ln \sigma_n}}{\sum_{i \in k_n} \frac{1}{\frac{2i-1}{2n+1}\cdot\frac{\pi}{2}}} = \frac{\sqrt{\ln \sigma_n}}{1 + \frac{1}{5} + \frac{1}{9} + \cdots + \frac{1}{2\sigma_{n-1}}} \cdot \frac{\pi}{2},$$

whence

$$\lim_{n\to\infty} \lambda_n = 0. \tag{27}$$

Having done all this we may pass to the construction of the required function $f(x)$.

Put $n_1 = 3$, and let n_2 denote an integer so large that it makes

$$| T_{n_2}[\varphi_{n_1} ; \pi] | < 1 \tag{28}$$

and

$$x^{(n_2)}_{n_2+\sigma_{n_2}} + \delta_{n_2} < x^{(n_1)}_{n_1+1} - \delta_{n_1} . \tag{29}$$

The first of these conditions can be satisfied because $\varphi_{n_1}(\pi) = 0$ and because at the point $x = \pi$ the interpolation process for $\varphi_{n_1}(x)$ is convergent, so that

$$\lim_{n\to\infty} T_n [\varphi_{n_1} ; \pi] = 0.$$

Condition (29) can also be satisfied because of

$$\lim_{n\to\infty} x^{(n)}_{n+\sigma_n} = \lim_{n\to\infty} \frac{2\pi}{2n+1} (n + \sigma_n) = \pi$$

and

$$\lim_{n\to\infty} \delta_n = 0.$$

Having chosen n_2 we denote by n_3 an integer sufficiently large for which

$$| T_{n_3}[\varphi_{n_1} + \varphi_{n_2} ; \pi] | < 1$$

and

$$x^{(n_3)}_{n_3+\sigma_{n_3}} + \delta_{n_3} < x^{(n_2)}_{n_2+1} - \delta_{n_2} .$$

By continuing this process we construct an infinite sequence of integers

$$n_1 < n_2 < n_3 < \cdots$$

such that

$$| T_{n_k} [\varphi_{n_1} + \varphi_{n_2} + \cdots + \varphi_{n_{k-1}} ; \pi] | < 1 \tag{30}$$

and

$$x^{(n_k)}_{n_k+\sigma_{n_k}} + \delta_{n_k} < x^{(n_{k-1})}_{n_{k-1}+1} - \delta_{n_{k-1}} . \tag{31}$$

By virtue of condition (31) no two of the functions

$$\varphi_{n_1}(x), \quad \varphi_{n_2}(x), \quad \varphi_{n_3}(x), \quad \cdots$$

can be different from zero simultaneously at the same point x. Thus, by the inequality

$$0 \leqq \varphi_n(x) \leqq \lambda_n$$

and the relation (27), the series

$$\varphi_{n_1}(x) + \varphi_{n_2}(x) + \varphi_{n_3}(x) + \cdots$$

converges *uniformly*. Hence we may set up the function

$$f(x) = \sum_{k=1}^{\infty} \varphi_{n_k}(x). \tag{32}$$

The function $f(x)$ is continuous. As all terms of the sum (32) are non-negative and vanish for $0 \leqq x \leqq \pi$ we have

$$\int_0^{2\pi} \left| \frac{f(x) - f(\pi)}{x - \pi} \right| dx = \int_\pi^{2\pi} \frac{f(x)}{x - \pi}\, dx = \sum_{k=1}^{\infty} \int_\pi^{2\pi} \frac{\varphi_{n_k}(x)}{x - \pi}\, dx,$$

and by (25)

$$\int_0^{2\pi} \left| \frac{f(x) - f(\pi)}{x - \pi} \right| dx < 1.$$

We also have:

$$T_{n_k}[f; \pi] = T_{n_k}[\varphi_{n_1} + \cdots + \varphi_{n_{k-1}}\, ; \pi] + T_{n_k}[\varphi_{n_k}\, ; \pi] + T_{n_k}[\varphi_{n_{k+1}} + \varphi_{n_{k+2}} + \cdots ; \pi].$$

Each of the functions $\varphi_{n_{k+1}}(x), \varphi_{n_{k+2}}(x), \cdots$, as well as their sum, becomes zero at the knots $x_i^{(n_k)}$ by virtue of (31).

Hence

$$T_{n_k}[\varphi_{n_{k+1}} + \varphi_{n_{k+2}} + \cdots ; \pi] = 0.$$

On combining this with (26) and (30) we get

$$T_{n_k}[f; \pi] > \sqrt{\ln \delta_{n_k}} - 1,$$

from which it follows that

$$\overline{\lim_{n \to \infty}}\, T_n[f; \pi] = +\infty$$

and at the point $x = \pi$ the process of interpolation diverges.

§5. An analogue to Abel-Poisson summation

By applying the identity

$$\frac{\sin \frac{2n+1}{2}}{2 \sin \frac{\alpha}{2}} = \tfrac{1}{2} + \cos \alpha + \cdots + \cos n\alpha$$

the interpolation polynomial may be written in the form

$$T_n(x) = \frac{a_0^{(n)}}{2} + \sum_{\nu=1}^{n} (a_\nu^{(n)} \cos \nu x + b_\nu^{(n)} \sin \nu x)$$

where

$$a_\nu^{(n)} = \frac{2}{2n+1} \sum_{k=0}^{2n} f(x_k^{(n)}) \cos \nu x_k^{(n)} \qquad (\nu = 0, 1, \cdots, n)$$

$$b_\nu^{(n)} = \frac{2}{2n+1} \sum_{k=0}^{2n} f(x_k^{(n)}) \sin \nu x_k^{(n)} \qquad (\nu = 1, 2, \cdots, n).$$

Put

$$T_n(r, x) = \frac{a_0^{(n)}}{2} + \sum_{\nu=1}^{n} r^\nu (a_\nu^{(n)} \cos \nu x + b_\nu^{(n)} \sin \nu x).$$

It is clear that

$$\lim_{n\to\infty} [\lim_{r\to 1} T_n(r, x)] = \lim_{n\to\infty} T_n(x).$$

Thus, if we take limits in this order the introduction of the polynomial $T_n(r, x)$ proves senseless.

In contrast to this, we may establish a theorem running as follows:

THEOREM 6. *If the function $f(x)$ is integrable (R), then at every point, where $f(x)$ is the derivative of its indefinite integral, we have*

$$\lim_{r\to 1-0} [\lim_{n\to\infty} T_n(r, x)] = f(x). \tag{33}$$

PROOF. Construct the Fourier-series for the function $f(x)$

$$\frac{a_0}{2} + \sum_{\nu=1}^{\infty} (a_\nu \cos \nu x + b_\nu \sin \nu x)$$

and set, for $0 < r < 1$,

$$T(r, x) = \frac{a_0}{2} + \sum_{\nu=1}^{\infty} r^\nu (a_\nu \cos \nu x + b_\nu \sin \nu x).$$

It is known that at the points x mentioned we have

$$\lim_{r\to 1-0} T(r, x) = f(x).$$

Thus in order to prove the theorem it is sufficient to show that

$$\lim_{n\to\infty} T_n(r, x) = T(r, x). \tag{34}$$

But this is semi-trivial. In fact, if ν is fixed, then

$$\lim_{n\to\infty} a_\nu^{(n)} = a_\nu, \qquad \lim_{n\to\infty} b_\nu^{(n)} = b_\nu \tag{35}$$

because $a_\nu^{(n)}$ and $b_\nu^{(n)}$ are the Riemann sums for integrals by which a_ν and b_ν are expressed.

On the other hand

$$| a_\nu^{(n)} | \leqq 2M, \qquad | b_\nu^{(n)} | \leqq 2M, \tag{36}$$

where M is the upper bound of $| f(x) |$. Putting

$$U_\nu = r^\nu(a_\nu \cos \nu x + b_\nu \sin \nu x) \qquad (\nu = 1, 2, 3, \cdots)$$

$$u_\nu(n) = \begin{cases} r^\nu(a_\nu^{(n)} \cos \nu x + b_\nu^{(n)} \sin \nu x) & (\nu = 1, 2, \cdots, n) \\ 0 & (\nu = n + 1, n + 2, \cdots) \end{cases}$$

$$U_0 = \frac{a_0}{2}, \qquad u_0(n) = \frac{a_0^{(n)}}{2}$$

we obtain

$$T(r, x) = \sum_{\nu=0}^{\infty} U_\nu, \qquad T_n(r, x) = \sum_{\nu=0}^{\infty} u_\nu(n).$$

Every term of the series $T_n(r, x)$ tends with $n \to \infty$ towards the corresponding term of the series $T(r, x)$ and by (36) the convergence of $T_n(r, x)$ is uniform with respect to the parameter n. This implies (34) and the theorem is proved.

LENINGRAD-BARNAUL.

ANNALS OF MATHEMATICS
Vol. 45, No. 3, July, 1944

REDUCTION OF THE SINGULARITIES OF ALGEBRAIC THREE DIMENSIONAL VARIETIES

BY OSCAR ZARISKI

(Received January 1, 1944)

CONTENTS

INTRODUCTION

1. In the present paper we give a solution of the problem of reduction of singularities of algebraic three-dimensional varieties.* In contrast with the

* The solution was first announced by the author in a conference report at the 1941 Summer Meeting of the American Mathematical Society in Chicago. See Bull. Amer. Math. Soc., vol. 48, n. 3 (1942), pp. 172–173.

two-dimensional case which has been studied by many authors, the three-dimensional problem is totally unexplored territory. As a challenge to the investigator it is second only to the general problem of the reduction of singularities of varieties of arbitrary dimension. How much more difficult is the general problem is of course impossible to say with certainty and precision at the present moment. We are inclined to conjecture that the difficulties in the general case and in the three-dimensional case are of comparable order of magnitude. However that may be, the three-dimensional problem offers an excellent testing ground for ideas and methods in relation to the general case, and for this reason its solution is presented here as a possible contribution to the general problem of reduction of singularities. We shall now endeavor to outline the main ideas of our solution.

2. By the theorem of local uniformization (see [6], p. 858), given any zero-dimensional valuation v of a field $\sum$ of algebraic functions of any number of variables (over a ground field of characteristic zero), there exists a projective model V of $\sum$ on which the center of v is a simple point. We shall show in a separate paper** that this theorem implies the existence of *a finite resolving system* for $\sum$, i.e. of a finite number of projective models of $\sum$ such that any zero-dimensional valuation v of $\sum$ has a simple center on at least one of the models (for the classical case, see our proof in [6], p. 854; compare also with footnote[15], section 25). We have shown in [7] that, in virtue of the existence of finite resolving systems, the solution of the problem of the reduction of singularities hinges entirely on the proof of the "fundamental theorem" stated in Part IV, section 25. Now the assumption of that theorem is that two given projective models V and V' constitute a resolving system for a given set N of zero-dimensional valuations. One is naturally led to consider tentatively the join V^* of V and V' ([8], p. 516). In general, V^* will not be a resolving model for N. However, the following condition is sufficient in order that V^* be a resolving model for N:

If v is any valuation in N and if P and P' are the centers of v on V and V' respectively, then one of the points P, P' is simple and at the same time is not fundamental for the birational correspondence T between V and V' (i.e. its quotient ring contains the quotient ring of the second point).

The gist of the considerations developed in the last section of this paper (section 25) is that in the three-dimensional case it is possible to transform V and V', by successive quadratic and monoidal transformations, into another pair of varieties for which the above condition is satisfied. The centers of these transformations are always selected as follows: if we are dealing with a simple point P of V which is an *isolated* fundamental point of T, then we apply to V a quadratic transformation of center P; if P lies on a simple fundamental curve Δ of T then we use a monoidal transformation of center Δ. In the second case it is, however, necessary to assume that *P is a simple point of* Δ, for otherwise the monoidal

** "The compactness of the Riemann manifold of an abstract field of algebraic functions", in course of publication in Bull. Amer. Math. Soc.

transformation will not necessarily satisfy the requirement of transforming the simple point P into simple points (section 5, a)). For this reason it is necessary first to resolve the singularities of Δ before a monoidal transformation is applied (section 20). Similar considerations apply to V' and to the successive transforms of V and V'. Given any valuation v in N, with centers P and P' on V and V' respectively, and if, say, P is the simple center, the intended effect of the successive transformations applied to V is *to augment the quotient ring of* P, so that ultimately P is replaced by a simple point whose quotient ring contains the quotient ring of P'. In this procedure we come up against the following complication (see case C, section 25): when we apply a monoidal transformation whose center is a *simple* curve Δ, we automatically augment the quotient ring of *each* point of Δ, hence also of such points of Δ which are *singular* for V and occur as centers of valuations in N. A lemma proved in section 24 enables us to overcome this difficulty.

3. When all incidental questions are discounted, there remains in the proof of the fundamental theorem a central idea, and that is the elimination, by successive quadratic and monoidal transformations, of the simple fundamental curves and of the isolated simple fundamental points of a given birational transformation T of a three-dimensional variety V into another variety V'. That this elimination is always possible is precisely what is asserted in Theorem 7 (section 19), or in the essentially equivalent Theorem 7′ (section 19). In this last theorem the elimination of the simple fundamental locus of T is presented as a process of elimination of the simple base locus (base curves and isolated base points outside the singular locus of V) of the linear system $|F|$ of surfaces on V which corresponds to the system of hyperplane sections of V'. *The exigencies of the proof of Theorem 7′ determine to a considerable extent the contents of the rest of the paper.* The scope of this theorem will be best understood if we analyze one implication of the theorem.

By a well-known theorem of Bertini, the singular points of the variable surface in the linear system $|F|$, outside the singular locus of V, lie necessarily on the base manifold of $|F|$. This theorem of Bertini is re-stated in algebraic terms, more appropriate for our purpose, in section 23, and we prove it for abstract varieties in a separate paper [9]. In virtue of this theorem of Bertini, any birational transformation of V which brings about the elimination of the simple base locus of the linear system $|F|$ *automatically reduces the singularities of the variable surface of the system* (not counting the singularities which fall at singular points of V). We render the term "variable surface" more precise by interpreting the linear system of *surfaces* $|F|$ as *one* surface F/K, over a suitable extended ground field K, embedded in the variety V/K. This way of looking at the linear system $|F|$ is part of our formulation of the theorem of Bertini. We see therefore that *Theorem 7′ implies a reduction theorem for algebraic surfaces,* namely *that it is possible to reduce the singularities of a surface F/K embedded in a three-dimensional variety V/K by successive quadratic and monoidal transformations of the ambient variety V/K* (Theorem 6, section 17). A geometric proof

of this theorem in the classical case, for surfaces embedded in a linear space, was presented by Beppo Levi in [2]. In Part III we prove this theorem for abstract surfaces. Here we make use of a local reduction theorem (Theorem 5, section 10) proved in Part II.

With the aid of the above reduction theorem, the proof of Theorem 7′ is carried out smoothly in two steps. First the theorem is proved directly for linear systems $| F |$ which, outside of the singular manifold of V, are free from singular base points (i.e., base points which are singular for each surface in the system). This is done in sections 20, 21, and 22. For arbitrary linear systems $| F |$ the proof is completed in section 23 by reducing the base singularities of $| F |$.

4. Our solution of the three-dimensional problem is complete in the case of ground fields of characteristic zero. However, we can point out with precision what has still to be done in order that the proof be complete also for fields of characteristic $p \neq 0$. Throughout the paper we have used the assumption that the field is of characteristic zero only in two instances: a) in Part IV, where we presuppose the local uniformization theorem for three-dimensional varieties; b) in sections 13, 14 and 15, where we prove the local reduction theorem for valuations of rank 1. These then are the only two gaps which would have to be filled in in order that our solution be valid for arbitrary ground fields of characteristic p. We must state explicitly that these gaps remain *even if the ground field is perfect.*

When we speak of a possible generalization of our solution of the three-dimensional problem to arbitrary fields of characteristic $p \neq 0$, we have in mind the following definition of a simple point: *a point P of an r-dimensional algebraic variety is simple if the ideal of non-units in the quotient ring of P has a base of r elements.* The elements of such a base are called *uniformizing parameters of P.* A simple point, according to this definition, is therefore characterized by the intrinsic condition that its quotient ring be a p-series ring in the sense of Krull [1]. It was pointed out to me by Chevalley and André Weil that this definition is in general not equivalent with the following classical definition: *if $(f_1, f_2, \cdots, f_m)$ is a base for the prime polynomial ideal in $k[x_1, x_2, \cdots, x_n]$ which defines the variety V, then a point P of V is simple if and only if the Jacobian matrix $\| \partial f_i / \partial x_j \|$ is of rank $n - r$ at P.* Simple points, according to this classical definition, remain simple upon an arbitrary extension of the ground field. For such points the term "absolutely simple" has therefore been proposed by Weil. An absolutely simple point P is necessarily a simple point, but if the residue field of P is not separable over the ground field the converse is not necessarily true. In a separate paper we shall analyse the relationship between simple and absolutely simple points, and we shall also characterize simple points by an algebraic condition involving the rank of a certain Jacobian matrix.

It is clear that if we mean by a non-singular variety one whose points are all absolutely simple, then not every field of algebraic functions has a non-singular model. For instance, for the existence of such a model it is at least necessary (but not sufficient) that the function field be separably generated over the

ground field. On the other hand, we conjecture that *every* field of algebraic functions (in any number of variables) possesses a model which is non-singular in the sense of our definition of simple points. It is in this sense that we speak, for instance, of the solution of the three-dimensional reduction problem for arbitrary ground fields of characteristic p.

One more remark. We make frequent use of the theorem that *every minimal prime ideal in the quotient ring of a simple point is a principal ideal.* In the case of characteristic zero and more generally for absolutely simple points, this theorem will be proved in our paper to be printed elsewhere (for algebraic surface, see our proof in [5], p. 664, Theorem 3). However, this theorem follows in all its generality, i.e. for arbitrary ground fields, from an unpublished result concerning p-series rings, due to I. Cohen. This theorem of Cohen states that *any p-series ring, whose characteristic coincides with the characteristic of its residue field K, can be embedded in a full power series ring over this field K* [see abstract of paper "Some theorems on local rings", by I. S. Cohen, Bull. Amer. Math. Soc., vol. 49, n. 1 (1943), p. 37].

Part I

Quadratic and monoidal transformations of surfaces embedded in a V_3

1. Multiple points of embedded surfaces

The object of our considerations in Parts I, II and III shall be an irreducible 3-dimensional algebraic variety V and a fixed irreducible algebraic surface F on V. The variety V may have singularities. *However, we stipulate that from now on whenever points or curves of V are considered it will be tacitly understood that these points or curves are simple for V.* To this we add that in the case of a simple curve of V it is not excluded that some special points of the curve may be singular for V, but if a point of the curve is mentioned it shall be understood that it is a simple point of V.

On the other hand, no such restrictions shall be imposed on the points and curves of the surface F. On the contrary, we shall be primarily interested in those points or curves of F which are singular for F (but are simple for V).

Quite generally, if H and G are irreducible algebraic varieties and if $H \subset G$, we shall denote by $Q_G(H)$ the quotient ring of H, regarded as a subvariety of G.

If H is simple for G and if say H and G are of dimension h and g respectively, then it is known that the ideal of non units in $Q_G(H)$ has a (minimal) base of $g - h$ elements. The elements $t_1, t_2, \cdots, t_{g-h}$ of any such minimal base shall be referred to as *uniformizing parameters* of $H(G)$. The notations $H(G)$ and $Q_G(H)$ are dictated by the necessity of avoiding confusion in cases when H occurs as a subvariety of more than one variety G. In all other cases the letter G will be omitted.

Let H be a point P or a curve Δ on F, and let $\mathfrak{m}$ be the ideal of non units in $\mathfrak{Z} = Q_V(H)$. The surface F is defined in $\mathfrak{Z}$ by a minimal prime ideal which is a principal ideal, say $\mathfrak{Z}\cdot\omega$, for H is simple for V (see Introduction).

DEFINITION 1. *H is ν-fold for the surface F if* $\omega \equiv 0(\mathfrak{m}^\nu)$, $\omega \not\equiv 0(\mathfrak{m}^{\nu+1})$.

This definition implies that ω can be expressed as a form f of degree ν in the uniformizing parameters t_i of $H(V)$, whose coefficients are in $\mathfrak{J}$ and do not all belong to $\mathfrak{m}$. Thus, if H is a point P, then

$$\omega = f(t_1, t_2, t_3), \tag{1}$$

and if H is a curve Δ, then

$$\omega = f(t_1, t_2). \tag{1'}$$

We shall also denote the multiplicity ν by $m_F(H)$.

LEMMA 1.1 *If $P \subset \Delta \subset F$, then $m_F(P) \geqq m_F(\Delta)$.* This is well known in the classical case. If the ground field is of characteristic zero or is a perfect field of characteristic p, the lemma can be proved by methods of formal differentiation (see Part II, section 13). However, since we shall give elsewhere (compare with Introduction, 4), a proof of Lemma 1.1 in the most general case, we shall omit the proof here.

2. Quadratic transformations

We apply to V a quadratic transformation T_1 whose center is a point P on F and we denote by V' and F' respectively the transforms of V and of F under T_1. Here F' is an irreducible surface on V' which is birationally equivalent to F. We denote by T the quadratic transformation from F to F' which is induced by T_1. We shall now state several salient features of T which we shall need in the sequel and which we have established elsewhere (see [8], p. 532).

By T_1 the point P is transformed into a surface $\Phi' = T_1[P]$.

a) *The surface Φ' is irreducible and each point of Φ' is simple for V'.*

Let $\mathfrak{J} = Q_V(P)$ and let $\mathfrak{m}$ be the ideal of non units in $\mathfrak{J}$. We denote by K the residue field of P, i.e. the field $\mathfrak{J}/\mathfrak{m}$. Let $S = S_2^K$ be the projective plane over K.

b) *The surface Φ' is in regular*[1] *birational correspondence with S.*

Let t_1, t_2, t_3 be a fixed set of uniformizing parameters of $P(V)$. If $f(z_1, z_2, z_3)$ is a form with coefficients in $\mathfrak{J}$, where z_1, z_2, z_3 are indeterminates, then we denote by $F(z_1, z_2, z_3)$ the form in $K[z_1, z_2, z_3]$ whose coefficients are the P-residues (i.e. the residues mod. $\mathfrak{m}$) of the coefficients of f. If $\xi \in \mathfrak{J}$, $\xi \equiv 0(\mathfrak{m}^\rho)$, $\xi \not\equiv 0(\mathfrak{m}^{\rho+1})$, we can write $\xi = f(t_1, t_2, t_3)$, where f is a form of degree ρ and where $F(z_1, z_2, z_3) \neq 0$. In this case we shall call $F(z_1, z_2, z_3)$ the *leading form* of ξ (see Krull [1], p. 207); it is independent of the particular way of representing ξ as a form of degree ρ in t_1, t_2, t_3. We identify the indeterminates z_1, z_2, z_3 with the homogeneous coördinates in S.

Let now H' be a point or an irreducible curve on Φ', or Φ' itself, and let $\bar{H}$ be the corresponding locus in S [by b)].

c) *The quotient ring $\mathfrak{J}' = Q_{V'}(H')$ consists of all quotients $f(t_1, t_2, t_3)/g(t_1, t_2, t_3)$,*

[1] For the definition of a regular correspondence see our paper [8], p. 513.

where $f(z_1, z_2, z_3)$, $g(z_1, z_2, z_3)$ *are forms of like degree, with coefficients in* $\mathfrak{J}$, *and where* $G(z_1, z_2, z_3) \neq 0$ *on* $\bar{H}$. *In other words,* $f(t)/g(t) \in Q_{V'}(H')$ *if and only if* $F(z)/G(z) \in Q_S(\bar{H})$.

From c), in the particular case $H' = \Phi'$, it follows that $Q_{V'}(\Phi')$ consists of all quotients $f(t)/g(t)$ such that $G(z) \neq 0$. In other words: if ξ and η are elements of $\mathfrak{J}$, then $\xi/\eta \in Q_{V'}(\Phi')$ if and only if the degree of the leading form of η is not greater than the degree of the leading form of ξ. Let μ be the degree of the leading form of η. If ξ/η is a non unit in $Q_{V'}(\Phi')$, then the degree of the leading form of ξ must be $\geqq \mu + 1$, and conversely. In this case we have: $\xi/t_1^{\mu+1} \in Q_{V'}(\Phi')$, $t_1^{\mu}/\eta \in Q_{V'}(\Phi')$, *whence* ξ/η *is divisible by* t_1 *in* $Q_{V'}(\Phi')$. Conversely, if this condition is satisfied, then ξ/η is a non unit in $Q_{V'}(\Phi')$. From this we conclude, on account of c), as follows:

d) *If* $z_1 \neq 0$ *on* $\bar{H}$, *the surface* Φ' *is defined in* $\mathfrak{J}'$ *by the principal ideal* $\mathfrak{J}' \cdot t_1$, *and we have:*

$$Q_S(\bar{H}) = \mathfrak{J}'/\mathfrak{J}' \cdot t_1; \qquad \mathfrak{J}' \sim Q_S(\bar{H}). \tag{2}$$

Let $\{\bar{t}_2, \bar{t}_3\}$ or $\{\bar{t}_2\}$ be a set of uniformizing parameters of $\bar{H}(S)$ according as H' (and therefore also $\bar{H}$) is a point or a curve. Let t_i' be an element of $\mathfrak{J}'$ which is mapped into $\bar{t}_i$ by the homomorphism (2). We put $t_1' = t_1$.

e) *According as* H' *is a point or a curve, the elements* t_1', t_2', t_3' *or the elements* t_1', t_2' *are uniformizing parameters of* $H'(V')$.

3. The effect of a quadratic transformation on the multiple points of F

We now turn our attention on the surfaces F and F' and on the quadratic transformation T, of center P, which T_1 induces between these two surfaces. To the fundamental point P there will correspond on F' an algebraic curve $T[P]$ which may be reducible. It is clear that this curve is the intersection of F' and Φ' and that [see (1)]

$$F(z_1, z_2, z_3) = 0 \tag{3}$$

is the equation of the plane curve in S which corresponds to $T[P]$ in the regular birational correspondence between Φ' and S. This plane curve is of order ν, where $\nu = m_F(P)$.

We take for H' [see c)] a point P' of $T[P]$ and we denote by $\bar{P}$ the point of S which corresponds to P'. We put $\omega' = \omega/t_1^{\nu}$.

LEMMA 3.1. *If* $z_1 \neq 0$ *at* $\bar{P}$, *then the principal ideal* $\mathfrak{J}' \cdot \omega'$ *is prime and defines the irreducible surface* F'.[2]

PROOF. We have to show that: 1) ω' is zero on F' and 2) that if an element ζ' of $\mathfrak{J}'$ is zero on F' then $\zeta' \in \mathfrak{J}' \cdot \omega'$. The first assertion is obvious, since $\omega = 0$ on F and $t_1 \neq 0$ on F (in view of our assumption that $z_1 \neq 0$ at $\bar{P}$). To prove 2), let $\zeta' = \xi/\eta$, where $\xi, \eta \in \mathfrak{J}$. Since $\zeta' = 0$ on F, ξ must be zero on F,

[2] Since P' is a point of F', the surface F' is given in $\mathfrak{J}'$ $(= Q_{V'}(P'))$ by a prime minimal ideal. When we say that $\mathfrak{J}' \cdot \omega'$ defines the surface F' we mean that this prime ideal coincides with $\mathfrak{J}' \cdot \omega'$.

whence $\xi = \omega \cdot \xi_1$, $\xi_1 \in \mathfrak{J}$. If ρ is the degree of the leading form of η, then the degree of the leading form of ξ is at least ρ [c)] and hence the degree of the leading form of ξ_1 is at least $\rho - \nu$. By c) we may assume that the leading form of η does not vanish at $\overline{P}$. Under this assumption the quotient $\xi_1 t_1'^{\nu}/\eta$ is in $\mathfrak{J}'$, whence $\zeta'/\omega' = \xi_1 t_1'^{\nu}/\eta \in \mathfrak{J}'$. This completes the proof.

LEMMA 3.2. *If P' is a point of $T[P]$, then $m_{F'}(P') \leqq m_F(P)$.*

PROOF. Let $\mathfrak{m}'$ and $\overline{\mathfrak{m}}$ denote the ideals of non-units in $\mathfrak{J}'$ and in $\overline{\mathfrak{J}} = Q_S(\overline{P})$ respectively, so that [see e)] $\mathfrak{m}' = \mathfrak{J}' \cdot (t_1', t_2', t_3')$ and $\overline{\mathfrak{m}} = \overline{\mathfrak{J}} \cdot (\bar{t}_2, \bar{t}_3)$. In the homomorphism $\mathfrak{J}' \sim \overline{\mathfrak{J}} = \mathfrak{J}'/\mathfrak{J}' \cdot t_1$ [see (2)], the ideal $\mathfrak{m}'$ is mapped into the ideal $\overline{\mathfrak{m}}$. The element ω' is mapped into the element $\bar{\omega} = F(z_1, z_2, z_3)/z_1^{\nu}$. If $\nu' = m_{F'}(P')$, then we must have, by Lemma 3.1, $\omega' \equiv 0(\mathfrak{m}'^{\nu'})$. Hence $\bar{\omega} \equiv 0(\overline{\mathfrak{m}}^{\nu'})$, i.e. *the point $\overline{P}$ must be a ν'-fold point of the curve* (3). Our lemma is therefore equivalent to the assertion that the plane curve (3), *which is of order* ν, cannot possess points of multiplicity $> \nu$. This assertion is trivial if the ground field k is algebraically closed. On the other hand, it is obvious from our definition 1 that the multiplicity of P for the curve $F(z) = 0$ does not diminish if we extend the ground field k. This establishes the lemma.

THEOREM 1. *If $T[P]$ possesses an irreducible component which is ν-fold for F' then $T[P]$ itself is irreducible and is a rational curve free from singularities.*

PROOF. Let Γ' be an irreducible component of $T[P]$ and let us assume that Γ' is ν-fold for F'. Let $\overline{\Gamma}$ be the irreducible plane curve in S which corresponds to Γ' in the regular birational correspondence between Φ' and S. The proof of Lemma 3.2 shows that $\overline{\Gamma}$ must be a ν-fold component of the curve (3). Since this curve is itself of order ν, it follows that the form $F(z_1, z_2, z_3)$ is necessarily the ν^{th} power of a linear form, i.e. the plane curve in S into which $T[P]$ is transformed must be a straight line, q.e.d.

4. Auxiliary considerations concerning ruled surfaces

Let Δ be an irreducible algebraic curve in an n-dimensional projective space S_n^k and let $\eta_0, \eta_1, \cdots, \eta_n$ be the homogeneous coördinates of the general point of Δ. If z_1, z_2 are algebraically independent over the field $k(\eta_0, \eta_1, \cdots, \eta_n)$, we consider the ruled surface R over k whose general point has the following homogeneous coördinates:

$$\rho\eta_{ij} = \eta_i z_j, \qquad i = 0, 1, 2, \cdots, n; \quad j = 1, 2,$$

where ρ is a factor of proportionality. This surface lies in a projective space of dimension $2n + 1$. Assuming that $\eta_0 \neq 0$ let $\xi_i = \eta_i/\eta_0$, $i = 1, 2, \cdots, n$, and let K be the field of rational functions on Δ: $K = k(\xi_1, \xi_2, \cdots, \xi_n)$. For a suitable choice of the factor ρ we can write: $\rho\eta_{0j} = z_j$, $\rho\eta_{ij} = \xi_i z_j$, and this shows that if K is taken as ground field then R is a straight line L and that the field of rational functions on R (over k) is the field $K(x)$, where $x = z_1/z_2$. Let $P(\alpha_0, \alpha_1, \cdots, \alpha_n)$ be a point of Δ, where the ratios of the α's are algebraic over k, and let z_1, z_2 be still regarded as algebraically independent over the residue field k^* of P. The irreducible curve, over k, which is defined by the general point

(η^*_{ij}), where $\eta^*_{ij} = \alpha_i z_j$, is a generator of the ruled surface R. We shall denote this generator by g_P. It is clear that g_P is a straight line l^* over k^*.

LEMMA 4.1. *The generator g_P is simple for R if and only if the point P is simple for Δ. If t is a uniformizing parameter of $P(\Delta)$ then t is also a uniformizing parameter of $g_P(R)$.*

PROOF. Let $\mathfrak{o} = Q_\Delta(P)$, $\mathfrak{o}_1 = Q_R(g_P)$. Any element of $\mathfrak{o}_1$ is of the form $f(\eta_{ij})/g(\eta_{ij})$, where f and g are forms of like degree with coefficient in k and where $g(\eta^*_{ij}) \neq 0$. If for any polynomial $F(x)$ in $\mathfrak{o}[x]$ we agree to denote by $F^*[x]$ the polynomial in $k^*[x]$ whose coefficients are the P-residues of the coefficients of $F(x)$, then we can say that $\mathfrak{o}_1$ consists of all quotients $F(x)/G(x)$, $F(x)$ and $G(x)$ in $\mathfrak{o}[x]$ and $G^*(x) \neq 0$.

Suppose that P is a simple point of Δ, with uniformizing parameter t. If $\omega_1 = F(x)/G(x) \in \mathfrak{o}_1$, then ω_1 is a non unit in $\mathfrak{o}_1$ if and only if $F^*(x) = 0$, i.e. if and only if each coefficient of $F(x)$ is a multiple of t (in $\mathfrak{o}$). Hence each non unit of $\mathfrak{o}_1$ is divisible by t (in $\mathfrak{o}_1$), and since t is itself obviously a non unit in $\mathfrak{o}_1$ it follows that g_P is simple for R and that t is a uniformizing parameter of $g_P(R)$.

Conversely, assume that g_P is simple for R and let t be a uniformizing parameter of $g_P(R)$. Without loss of generality we may assume that t is a polynomial in $\mathfrak{o}[x]$, say

$$t = t_0x^m + t_1x^{m-1} + \cdots + t_m, \qquad t_i \in \mathfrak{o}.$$

Since $t = 0$ on g_P, it follows $t_i = 0$ at P, whence the coefficients t_i are non units in $\mathfrak{o}$. We can therefore write:

$$t_i \cdot G(x) = tF_i(x), \qquad G^*(x) \neq 0, \qquad i = 0, 1, \cdots, m. \tag{4}$$

Let now ω be an arbitrary non unit in $\mathfrak{o}$. Since ω is also a non unit in $\mathfrak{o}_1$, we have: $\omega \cdot B(x) = t \cdot A(x)$, $B(x)$ and $A(x)$ in $\mathfrak{o}[x]$ and $B^*(x) \neq 0$. If therefore b is any coefficient of $B(x)$, the product ωb belongs to the ideal $\mathfrak{o} \cdot (t_0, t_1, \cdots, t_m)$. In particular, if b is one of the coefficients of $B(x)$ which is not zero at P, then b is a unit in $\mathfrak{o}$, and we conclude that $\omega \in \mathfrak{o} \cdot (t_0, t_1, \cdots, t_m)$. Hence $\mathfrak{o} \cdot (t_0, t_1, \cdots, t_m)$ is the ideal $\mathfrak{m}$ of non units in $\mathfrak{o}$. Since $\mathfrak{m} \neq \mathfrak{m}^2$ it follows from (4) that we cannot have $F^*_i(x) = 0$ for all i. Let, say, $F^*_\nu(x) \neq 0$, and let c be a coefficient of $F_\nu(x)$ which is not zero at P and which is therefore a unit in $\mathfrak{o}$. We have $t_iF_\nu(x) = t_\nu F_i(x)$ and consequently $ct_i \equiv 0(\mathfrak{o} \cdot t_\nu)$. This shows that $\mathfrak{m} = \mathfrak{o} \cdot t_\nu$, q.e.d.

LEMMA 4.2. *If P is a simple point of Δ, then every point of g_P is simple for R.*

PROOF. The general point of g_P is (η^*_{ij}), where $\eta^*_{ij} = \alpha_i z_j$. Any point P^* of g_P has as homogeneous coördinates the products $\alpha_i\beta_j$, where the ratio β_1/β_2 is algebraic over k. The quotient ring $Q_R(P^*)$ consists of all quotients $F(z_1, z_2)/G(z_1, z_2)$, where F, G are forms of like degree with coefficients in $\mathfrak{o}$ and where the form $G^*(z_1, z_2)$ in $k^*[z_1, z_2]$ is such that $G^*(\beta_1, \beta_2) \neq 0$. The quotient ring $Q_{g_P}(P^*)$ consists of the quotients $F^*(z_1, z_2)/G^*(z_1, z_2)$. This shows in the first place that P^* is a simple point of g_P, and this is true even if P is singular for Δ. This is in agreement with a preceding remark to the effect that g_P is a straight

line. Let $\tau^* = F^*(z_1, z_2)/G^*(z_1, z_2)$ be a uniformizing parameter of $P^*(g_P)$ and—assuming that P is simple for Δ—let t be a uniformizing parameter of $P(\Delta)$. Then it is immediately seen that if we put $\tau = F(z_1, z_2)/G(z_1, z_2)$, then t and τ are uniformizing parameters of $P^*(R)$.

5. Monoidal transformations

We apply to V a monoidal transformation T_1 whose center is an irreducible curve Δ on F and we denote by V' and F' the transforms of V and F respectively. We collect in this section some fundamental properties of T_1. For the proofs see our paper [8].

Under T_1 the curve Δ is transformed into a surface, which may be reducible. However, only one irreducible component of this surface corresponds to Δ itself. We shall denote this component by Φ'. The remaining components correspond to special points of Δ, points which are either singular for V or for Δ, and do not interest us at present.

a) *The points of the surface Φ' which correspond to simple points*[3] *of Δ are simple for V' and for Φ'.*

We denote by $\mathfrak{J}$ the quotient ring $Q_V(\Delta)$ and by $\mathfrak{m}$ the ideal of non units in $\mathfrak{J}$. Let $\eta_0, \eta_1, \cdots, \eta_n$ be the homogeneous coördinates of the general point of Δ, and let $K = \mathfrak{J}/\mathfrak{m}$ be the field of rational functions on Δ. We introduce the ruled surface R relative to the curve Δ (see preceding section) and we denote by N' the subvariety of Φ' whose points correspond to the points of Δ which are singular either for Δ or for V.

b) *The surfaces Φ' and R are birationally equivalent, and corresponding points P' and P^* ($P' \epsilon \Phi'$, $P^* \epsilon R$) are such that if $T_1(P) = P'$ then $P^* \epsilon g_P$. This birational correspondence is regular at each point of $\Phi' - N'$.*

We shall denote by g'_P the transform of P on Φ'. By a), b) and by Lemmas 4.1 and 4.2 it follows that:

c) *If P is simple for Δ (and also for V), then g'_P is in regular birational correspondence with the generator g_P and is therefore an irreducible rational curve free from singularities; moreover, g'_P and each point of g'_P is simple for V' and for Φ'.*

In addition to the ground field k and to the field K of rational functions on Δ we shall have occasion to deal with the residue field of P. This last field shall be denoted by k^*. We shall use small Latin letters f, g, h, etc. to denote polynomials (in any number of variables) with coefficients in $\mathfrak{J}$. If the coefficients of such polynomials are reduced mod Δ, the resulting polynomials will have coefficients in K and shall be denoted by the corresponding capital Latin letters $F, G, H \cdots$. If the coefficients of $f, g, h \cdots$ are not only in $\mathfrak{J}[= Q_V(\Delta)]$ but also in $Q_V(P)$, then their coefficients can be reduced mod P, and the resulting polynomials, with coefficients in k^*, shall be denoted by $f^*, g^*, h^*, \cdots$ respectively.

Let t_1, t_2 be a fixed pair of uniformizing parameters of $\Delta(V)$ and let z_1, z_2

---

[3] But these points must also be simple for V (see section 1).

be indeterminates. If $\mathfrak{m}$ denotes the ideal of non units in $\mathfrak{J}$ and if ξ is an element of $\mathfrak{J}$ which is exactly divisible by $\mathfrak{m}^\rho$, then $\xi = f(t_1, t_2)$, where f is a form of degree ρ, and $F(z_1, z_2)$—the leading form of ξ—is not zero.

In the sequel only points and curves of Φ' are considered which do not belong to N'.

Let H' be an irreducible curve or a point on Φ', or Φ' itself. If H' is a curve, not in the set $\{g'_P\}$, or if $H' = \Phi'$, let H^* be the corresponding locus on R. Since H^* is not a generator g_P, no element of K, different from zero, is zero on H^*. Hence if R is regarded as a straight line over K, then H^* is a point of that line or the line itself. Consequently it makes sense to say that a form $F(z_1, z_2)$ does or does not vanish on H^*.

On the other hand, if H' is either a point of Φ' or if $H' = g'_P$, then H' corresponds to a point P of Δ. Then H^* is either a point P^* or a generator g_P. We select uniformizing parameters t_1, t_2, t_3 of $P(V)$ in such a fashion that t_1, t_2 be uniformizing parameters of $\Delta(V)$.

In both cases let H denote the corresponding locus in Δ ($H = \Delta$ in the first case, $H = P$ in the second case).

d) *The quotient ring* $\mathfrak{J}' = Q_{V'}(H')$ *consists of all quotients* $f(t_1, t_2)/g(t_1, t_2)$ *of forms of like degree, with coefficients in* $Q_V(H)$, *such that* $F(z_1, z_2)/G(z_1, z_2) \in Q_R(H^*)$.

In particular, if $H' = \Phi'$, then it follows that $Q_{V'}(\Phi')$ consists of the quotients ξ/η, ξ and η in $\mathfrak{J}$, such that the degree of the leading form of η is not greater than the degree of the leading form of ξ. From this it follows as in section 2 that:

e) *If* $z_1 \neq 0$ *on* H^* (*i.e. if* $\eta_{01}, \eta_{11}, \cdots, \eta_{n1}$ *are not all zero on* H^*), *then the surface* Φ' *is defined in* $\mathfrak{J}'$ *by the prime principal ideal* $\mathfrak{J}' \cdot t_1$, *and we have*

$$Q_R(H^*) \cong \mathfrak{J}'/\mathfrak{J}' \cdot t_1 . \tag{5}$$

A similar result holds if $z_2 \neq 0$ *on* H^*, *in which case* t_1 *is to be replaced by* t_2.

f) *The elements* t_1, t_3 *and also the elements* t_2, t_3 *are uniformizing parameters of* $g'_P(V')$. *If* $H' = P'$ *is a point of* g'_P *with* $H^* = P^*$ *as the corresponding point on* R, *and if* $z_1 \neq 0$ *at* P^*, *then* (t_1, t'_2, t_3) *are uniformizing parameters of* $P'(V')$, *where* t'_2 *is any element of* $\mathfrak{J}'$ *which in the homomorphism*[4] $\mathfrak{J}' \sim Q_{g_P}(P^*)$ *is mapped onto a uniformizing parameter of* $P^*(g_P)$.

6. The effect of a monoidal transformation on the multiple points of F

We now study the birational transformation T from F to F' induced by the monoidal transformation T_1. To the curve Δ there will correspond on F' an algebraic curve $T[\Delta]$, which may be reducible. The total intersection of F' and Φ' may contain, in addition to $T[\Delta]$, also a finite number of curves g'_P and a finite number of isolated points P'. These isolated points necessarily are

---

[4] We note that $\mathfrak{J}' \sim Q_R(H^*)$, by (5), and that if H^* is a point P^* on g_P then $Q_R(P^*) \sim Q_{g_P}(P^*)$.

singular for V' (and hence belong to N'), for at any common point of F' and Φ' which is simple for V' the intersection of F' and Φ' is locally pure 1-dimensional. As to the curves g'_P which may possibly occur in the total intersection of F' and Φ', we shall see later on that they necessarily correspond to certain special points P of Δ (see Lemmas 6.2 and 6.5).

Let H' be an irreducible component or a point of $T[\Delta]$ and let H and H^* denote respectively the locus on F and R which corresponds to H'. We denote by $\mathfrak{J}'$ the quotient ring $Q_{V'}(H')$. We put $\omega' = \omega/t_1^\nu$, where $\omega \in Q_V(H)$ [see (1) and (1')].

LEMMA 6.1. *If $z_1 \neq 0$ on H^*, then the principal ideal $\mathfrak{J}' \cdot \omega'$ is prime and defines the irreducible surface F'.*

The proof is the same as that of Lemma 3.1.

LEMMA 6.2. *If P is a point of Δ which is simple for Δ (and also for V) and if the curve g'_P lies on F', then necessarily $m_F(P) > m_F(\Delta)$.*

PROOF. If $\nu = m_F(\Delta)$ and if t_1, t_2, t_3 are uniformizing parameters of $P(V)$ such that t_1, t_2 are uniformizing parameters of $\Delta(V)$, then ω $[= f(t_1, t_2, t_3)]$ can be expressed as a form $g(t_1, t_2)$, of degree ν, with coefficients in the ring $Q_V(P)$. Assuming that $z_1 \neq 0$ on the generator g_P of R, it follows from d), section 5, that $g(t_1, t_2)/t_1^\nu \in Q_{V'}(g'_P)$. If we now assume that $g'_P \in F'$, then the element $g(t_1, t_2)/t_1^\nu$ must be a non-unit in $Q_{V'}(g'_P)$. That implies $g^*(z_1, z_2) = 0$, where the coefficients of g^* are the P-residues of the coefficients of g. Hence the coefficients of $g(t_1, t_2)$ are non units in $Q_V(P)$, whence ω can be written as a form in t_1, t_2, t_3, with coefficients in $Q_V(P)$, *of degree greater than* ν. Hence $m_F(P) > \nu$, q.e.d.

LEMMA 6.3. *If P is a simple point of Δ such that $m_F(P) = m_F(\Delta)$ and if P' is a point of F' which corresponds to P, then $m_{F'}(P') \leqq m_F(P)$.*

PROOF. Let $\nu = m_F(P) = m_F(\Delta)$, $\nu' = m_{F'}(P')$, and let P^* be the point of R which corresponds to P' [see b), section 5]. Let $\mathfrak{J}$, $\mathfrak{J}'$ and $\bar{\mathfrak{J}}$ denote respectively the quotient rings $Q_V(P)$, $Q_{V'}(P')$ and $Q_R(P^*)$, and let $\mathfrak{m}$, $\mathfrak{m}'$ and $\bar{\mathfrak{m}}$ denote the ideal of non units in these rings. Assuming that $z_1 \neq 0$ at P^*, we have:

$$\omega' = \omega/t_1^\nu = g(t_1, t_2)/t_1^\nu \equiv 0(\text{mod } \mathfrak{m}'^{\nu'}).$$

By the homomorphism $\mathfrak{J}' \sim \bar{\mathfrak{J}}$ [see (5)] the element ω' is mapped into the element $\bar{\omega} = G(z_1, z_2)/z_1^\nu$, and we have

$$\bar{\omega} = G(z_1, z_2)/z_1^\nu \equiv 0(\bar{\mathfrak{m}}^{\nu'}). \tag{6}$$

Let now $\mathfrak{J}^* = Q_{g_P}(P^*)$ and let $\mathfrak{m}^*$ be the ideal of non units in $\mathfrak{J}^*$. From (6) follows (see footnote[4]):

$$\omega^* = g^*(z_1, z_2)/z_1^\nu \equiv 0(\mathfrak{m}^{*\nu'}). \tag{7}$$

Since $m_F(P) = m_F(\Delta)$, $g^*(z_1, z_2)$ is not identically zero and is a form of degree ν with coefficients in k^* (k^* = residue field of P). Hence if we put $z_2/z_1 = x$, then ω^* is a polynomial in $k^*[x]$, of degree $\leqq \nu$, while $\mathfrak{J}^*$ is the quotient ring of a prime ideal in $k^*[x]$. Consequently, by (7), we conclude that $\nu' \leqq \nu$, q.e.d.

LEMMA 6.4. *If H' is an irreducible component of $T[\Delta]$, then $m_{F'}(H') \leqq m_F(\Delta)$.*

PROOF. In this case, if K is taken as ground field, then H^* is a *point* of R (over K) and the lemma follows, as in the preceding lemma, from the fact that $G(z_1, z_2)$ is a form of degree ν.

LEMMA 6.5. *If $\nu = m_F(\Delta)$, all but a finite number of points of Δ are exactly ν-fold for F.*

PROOF. The surface F is defined in $Q_V(\Delta)$ by a principal ideal (ω), where ω is a form $g(t_1, t_2)$ of degree ν with coefficients in $Q_V(\Delta)$. Here t_1, t_2 are uniformizing parameters of $\Delta(V)$, and we may suppose that they are polynomials in the non-homogeneous coördinates ξ_i of the general point of V. If $\mathfrak{o}$ denotes the ring $k[\xi_1, \xi_2, \cdots, \xi_n]$ (we assume that V is immersed in an S_n^k), we consider the ideal $\mathfrak{o}\cdot(t_1, t_2)$ and the points P of Δ satisfying the following conditions:

1) P is at finite distance (with respect to the above non-homogeneous coördinates ξ_i) and is simple for V and for Δ.

2) The coefficients of $g(t_1, t_2)$ are in $Q_V(P)$ and are not zero at P.

It is clear that the number of points P on Δ which do not satisfy both conditions 1) and 2), is finite. Now we show that if conditions 1) and 2) are satisfied, then $m_F(P) = \nu$.

Since P is simple for Δ, there exists an element t_3 in $Q_V(P)$ such that t_1, t_2, t_3 are uniformizing parameters of $P(V)$. But then, in view of condition 2, $g(t_1, t_2)$ is a form *in the uniformizing parameters* of P (although t_3 does not actually occur in the form), with coefficients in $Q_V(P)$ which are not all zero at P. Our lemma will therefore follow if we show that ω remains prime in $Q_V(P)$. Suppose that ω factors in $Q_V(P)$, say $\omega = \omega_1\omega_2$. Since ω is prime in $Q_V(\Delta)$, one of the factors, say ω_1, must be a unit in $Q_V(\Delta)$, and the other factor ω_2 must be divisble by the ideal $(t_1, t_2)^\nu$, in $Q_V(P)$. But then ω_1 must also be a unit in $Q_V(P)$, for otherwise all the coefficients of $g(t_1, t_2)$ would be zero at P.

THEOREM 2. *If $T[\Delta]$ possesses an irreducible component which is ν-fold for F', then $T[\Delta]$ itself is an irreducible curve, birationally equivalent to Δ, and the birational correspondence between Δ and $T[\Delta]$ is regular at each simple point P of Δ, provided $m_F(P) = m_F(\Delta)$.*

PROOF. Let Δ' be an irreducible component of $T[\Delta]$ which is ν-fold for F' and let Δ^* be the corresponding curve on R. The curves on R which correspond to the various components of $T[\Delta]$ (in the birational correspondence between Φ' and R), *regarded as points over K*, are given by the zeros of the form $G(z_1, z_2)$. This form in $K[z_1, z_2]$ is of degree ν. If Δ' is ν-fold for F', then Δ^* must correspond to a ν-fold root of $G(1, x)$, where $x = z_2/z_1$. Hence $G(z_1, z_2)$ must be the ν^{th} power of a linear form in $K[z_1, z_2]$, and this shows that $T[\Delta]$ itself is irreducible, $T[\Delta] = \Delta'$, and that the field of rational functions on Δ' coincides with K, i.e. Δ' is birationally equivalent to Δ.

Let now P and P' be corresponding points of Δ and of Δ', where we assume that $m_F(P) = m_F(\Delta)$ and that P is simple for Δ. To complete the proof of the theorem we have to show that $Q_\Delta(P) = Q_{\Delta'}(P')$. Since P is simple for Δ we may assume that the uniformizing parameters t_1, t_2 of $\Delta(V)$ together with some

element t_3 form a set of uniformizing parameters of $P(V)$. Then the coefficients of $g(t_1, t_2)$ are in $Q_V(P)$ and are not all zero at P, since $m_F(P) = m_F(\Delta)$ (see proof of Lemma 6.5). Therefore the coefficients of $G(z_1, z_2)$ are in $Q_\Delta(P)$ and are not all zero at P. By the first part of the proof we have:

$$G(z_1, z_2) = (a_2 z_1 - a_1 z_2)^\nu, \qquad a_1, a_2 \in K,$$

and now we know that a_1 and a_2 are both in $Q_\Delta(P)$ and are not both zero at P. Let, say, $a_2 \neq 0$ at P and let a_1^*, a_2^* be the P-residues of a and b. In the regular birational correspondence between Φ' and R let Δ^* be the irreducible curve on R which corresponds to Δ' and let P^* be the point of Δ^* which corresponds to P'. Then it is clear that $a_2 z_1 - a_1 z_2 = 0$ on Δ^* and that $z_1 : z_2 = a_1^* : a_2^*$ at P^*. The quotient ring $Q_R(P^*)$ consists of all quotients $A(z_1, z_2)/B(z_1, z_2)$ of forms of like degree, with coefficients in $Q_\Delta(P)$, such that $B(a_1^*, a_2^*) \neq 0$. Since $z_1 : z_2 = a_1 : a_2$ on Δ^*, it follows that $Q_{\Delta^*}(P^*)$ consists of all quotients $A(a_1, a_2)/B(a_1, a_2)$, such that $B(a_1^*, a_2^*) \neq 0$. That shows that $Q_\Delta(P) = Q_{\Delta^*}(P^*)$, and since the birational correspondence between Φ' and R is regular at P^*, we also have $Q_{\Delta^*}(P^*) = Q_{\Delta'}(P')$. This completes the proof.

Corollary. *If T transforms the ν-fold curve Δ of F into a ν-fold curve Δ' of F', then simple points of Δ which are ν-fold for F go into simple points of Δ'. In particular, if Δ is free from singularities and if no point of Δ is of higher multiplicity than ν for F, then Δ' is free from singularities.*

7. Further remarks concerning singular points of F whose multiplicity is not lowered by a quadratic or by a monoidal transformation

We have seen in the preceding section (Theorem 2) that if a ν-fold curve Δ of F is transformed by the monoidal transformation T of center Δ into a ν-fold curve Δ' of F', then the two curves Δ and Δ' are birationally equivalent. We can express this result by saying that:

a) *residue field of Δ = residue field of Δ'.*

Now consider the case of a ν-fold point P of F belonging to the ν-fold curve Δ, and let us assume that a point P' of F' which corresponds to P is also ν-fold for F'. From the proof of Lemma 6.3 we deduce that the form $g^*(z_1, z_2)$ must be the ν^{th} power of a linear form in $k^*[z_1, z_2]$, where k^* is the residue field of P. Hence if P^* denotes the point of R which corresponds to P', then the P^*-residue of the ratio z_2/z_1 is an element of k^*, since $g^*(z_1, z_2) = 0$ at P^*. This shows that the residue field of P^*, and hence also the residue field of P', coincides with the residue field of P. Moreover this also shows that P' is the only point of $T[\Delta]$ which corresponds to P. Hence

b) *residue field of P = residue field of P', and $T[P] = P'$.*

We now go back to the case of a quadratic transformation T, studied in section 3. We suppose that $T[P]$ carries a point P' which is ν-fold for F', where $\nu = m_F(P)$. This point P' must then correspond to a ν-fold point P of the curve (3) (see proof of Lemma 3.2). The conditions under which a plane curve of order ν can possess a ν-fold point are elicited in the following lemma:

LEMMA 7.1. *If $F(z_1, z_2, z_3)$ is a form of degree ν in $K[z_1, z_2, z_3]$, different from zero, and if the plane curve $F(z_1, z_2, z_3) = 0$ has a ν-fold point P, then the curve consists of ν (not necessarily distinct) lines through P.*

PROOF. We assume that $z_3 \neq 0$ at P and we use the non-homogeneous coördinates $x = z_1/z_3$, $y = z_2/z_3$. Let $F(x, y, 1) = f(x, y)$. Our lemma asserts that if the point P is ν-fold for the curve $f(x, y) = 0$, then the polynomial $f(x, y)$ is either the ν^{th} power of a linear polynomial in $K[x, y]$ or is a form of degree ν in two such linear polynomials which both vanish at P. Thus the lemma implies at any rate that there exists a linear polynomial which is zero at P. On the other hand, if we can show that there exists a linear polynomial which is zero at P, then the proof of the lemma can be readily completed as follows. Without loss of generality we may assume that $x = 0$ at P. Let S denote the (x, y)-plane and let l denote the line $x = 0$. We put $\mathfrak{o} = Q_S(P)$, $\mathfrak{o}_1 = Q_l(P)$ and we denote by $\mathfrak{m}$ and $\mathfrak{m}_1$ respectively the ideals of non-units in $\mathfrak{o}$ and $\mathfrak{o}_1$. We have by hypothesis: $f(x, y) \equiv 0(\mathfrak{m}^\nu)$. Hence if we put $f_0(y) = f(0, y)$, we also have: $f_0(y) \equiv 0(\mathfrak{m}_1^\nu)$. If $f_0(y) \not\equiv 0$, then the congruence $f_0(y) \equiv 0(\mathfrak{m}_1^\nu)$ is possible only if $f_0(y) = (c_0 y + c_1)^\nu$, $c_0, c_1 \in K$, where $c_0 y + c_1 = 0$ at P. Without loss of generality we may assume that $f_0(y) = y^\nu$. The point P is then the origin $x = y = 0$ and $f(x, y)$ is necessarily a form of degree ν in x and y. Suppose now that $f_0(y) = 0$. Then $f(x, y) = x f_1(x, y)$, and since $x \not\equiv 0(\mathfrak{m}^2)$, it follows that $f_1(x, y) \equiv 0(\mathfrak{m}^{\nu-1})$. Since f_1 is of degree $\nu - 1$, the proof is completed by induction with respect to ν.

In order to prove the existence of a linear polynomial in $K[x, y]$ which vanishes at P, we use formal differentiation. In order to include fields of arbitrary characteristic, we shall use the well-known operator:

$$D_{ij}^{(m)} = \frac{1}{i!\,j!} \cdot \frac{\partial^{(m)}}{\partial x^i \partial y^j}, \; i + j = m.$$

It is immediately seen that if $f(x, y) \equiv 0(\mathfrak{m}^\nu)$, then $D_{ij}^{(m)} f(x, y) \equiv 0(\mathfrak{m}^{\nu - m})$. Since $D_{ij}^{(m)} f(x, y)$ is of degree $\leqq \nu - m$, the existence of at least one linear polynomial which vanishes at P is assured (by induction on ν), provided not all the partial derivatives $D_{ij}^{(m)}$ of $f(x, y)$, $m = 1, 2, \cdots, \nu - 1$, vanish identically. In the contrary case we must have $\nu = p^e$ and f must be a polynomial of the form

$$f(x, y) = c_0 + c_1 x^{p^e} + c_2 y^{p^e}, \tag{8}$$

where we may assume that $c_2 \neq 0$.

Let $\bar{x}$, $\bar{y}$ be the coördinates of P in a suitable algebraic extension field of K. Let $g(x)$, $h(x, y)$ be polynomials in $K[x, y]$ defined by the following conditions: 1) the leading coefficient of $g(x)$ is 1 and also the leading coefficient of $h(x, y)$, regarded as polynomial in y with coefficients in $k[x]$, is 1; 2) $g(x)$ is irreducible in $K[x]$ and $g(\bar{x}) = 0$; 3) the degree of $h(x, y)$ in x is less than the degree of $g(x)$; 4) $h(\bar{x}, \bar{y}) = 0$ and $h(\bar{x}, y)$ is irreducible in $K(\bar{x})[y]$. Then it is known ([8], p. 541) that $g(x)$ and $h(x, y)$ are not only uniformizing parameters of P, but also form a base for the prime ideal of P in $K[x, y]$. If $g(x)$ is of degree 1, then

there is nothing to prove. We therefore assume that $g(x)$ is of degree greater than 1.

From our hypothesis that $f(x, y) \equiv 0(\mathfrak{m}^\nu)$ follows immediately that $f(\bar{x}, y)$ must be divisible by $h(\bar{x}, y)^\nu$ (in $k(\bar{x})[y]$). Since $f(\bar{x}, y) \neq 0$, it follows that $h(x, y)$ must be of degree 1 in y, say

$$h(x, y) = y + A(x).$$

Consider the following polynomial in $K[x]$:

$$B(x) = f(x, y) - c_2[y + A(x)]^{p^e}.$$

Since the polynomial on the right belongs to the ideal $\mathfrak{m}^\nu$ and since $B(x)$ depends only on x, we conclude that $B(x)$ is divisible by $[g(x)]^{p^e}$. Since the degree of $A(x)$ is less than the degree of $g(x)$ and since by hypothesis the degree of $g(x)$ is greater than 1, comparison of the degrees of $B(x)$ and of $[g(x)]^{p^e}$ shows that $B(x)$ must be identically zero. Hence

$$c_0 + c_1x^{p^e} + c_2y^{p^e} = c_2y^{p^e} + c_2[A(x)]^{p^e}.$$

From this we conclude that $A(x)$ is linear, say $A(x) = d_0x + d_1$. Hence, $h(x, y)$ is a linear polynomial which vanishes at P. This completes the proof of the lemma.

In applying this lemma to the ν-fold point P of the curve (3) we observe, in the case of the lemma, that if there exist two independent linear polynomials in $K[x, y]$ which vanish at P, then the coordinates $\bar{x}$, $\bar{y}$ of P are in K. Since the coefficients of the $F(z_1, z_2, z_3)$ in (3) are in the residue field K of P, we conclude with the following theorem:

THEOREM 3. *If P is a ν-fold point of F and if among the points of F' which correspond to P in the quadratic transformation of center P there is one, say P', which is also ν-fold for F', then either $T[P]$ is an irreducible rational curve free from singularities (and—in fact—is equivalent to a straight line over K under a regular birational transformation) or $T[P]$ consists of several such curves through P', and then the residue field of P' coincides with the residue field of P.*

In the sequel we shall find it convenient to use the following phraseology: we shall say that the multiplicity of the center H of a quadratic or of a monoidal transformation $T(H \subset F)$ is *uniformly reduced* by T, if for any two corresponding points P and P' of F and F' respectively, such that P is a simple point of H and $m_F(P) = m_F(H)$, it is true that $m_{F'}(P') < m_F(P)$. (In the case of a quadratic transformation P is H itself and the condition $m_F(P) = m_F(H)$ is vacuous). In the problem of the reduction of the singularities of F it is true that the reduction process suffers a setback whenever at some step of the process a quadratic or monoidal transformation T fails to reduce uniformly the multiplicity of its own center. However, the results a), b) and Theorem 3 of this section show that, with one exception (see Theorem 3), the setback is always compensated, to a certain degree, by *the invariance of the residue field* of the point or of the curve under consideration. This compensation is brought about by the fact that

when the residue field is invariant the transformation T operates *locally* very much in the same fashion as it would if the ground field k was algebraically closed. We proceed to develop this idea.

1) Let first T be a monoidal transformation of center Δ, and let us suppose that some irreducible component Δ' of $T[\Delta]$ has the same residue field as Δ (i.e. that Δ and Δ' are birationally equivalent). Let t_1 and t_2 be uniformizing parameters of $\Delta(V)$. By d), section 5, either t_2/t_1 or t_1/t_2 belongs to $Q_{V'}(\Delta')$. Let, say, $t_2/t_1 \epsilon Q_{V'}(\Delta')$. Since Δ and Δ' have the same residue field, there exists an element ω in $Q_V(\Delta)$ which has the same Δ'-residue as t_2/t_1, and hence $(t_2 - t_1\omega)/t_1$ is a non unit in $Q_{V'}(\Delta')$. It is clear that also t_1 and $t_2 - t_1\omega$ are uniformizing parameters of $\Delta(V)$. Hence we may replace t_2 by $t_2 - t_1\omega$ and we may therefore assume that t_2/t_1 is a non unit in $Q_{V'}(\Delta')$. Using the results of section 5, especially d), we now conclude as follows:

If we put

$$t'_1 = t_1, \qquad t'_2 = t_2/t_1, \tag{9}$$

then t'_1 and t'_2 are uniformizing parameters of $\Delta'(V')$, and the quotient ring $Q_{V'}(\Delta')$ consists of all quotients $\phi(t'_1, t'_2)/\psi(t'_1, t'_2)$, where ϕ and ψ are polynomials with coefficients in $Q_V(\Delta)$ and where the "constant term" of ψ is a unit in $Q_V(\Delta)$.

The equations (9) can be regarded *locally* (i.e. in regard to the pair of corresponding curves Δ and Δ') as the equations of T. The form of these equations is the same as that of a quadratic transformation between two planes (t_1, t_2) and (t'_1, t'_2), over an algebraically closed ground field.

2) The transformation T still being monoidal, with center Δ, let P be a simple point of Δ and let us suppose that a point P' of F' which corresponds to P is such that its residue field is the same as that of P. Let t_1, t_2, t_3 be uniformizing parameters of $P(V)$ such that t_1, t_2 are uniformizing parameters of $\Delta(V)$. By d), section 5, we may assume that $t_2/t_1 \epsilon Q_{V'}(P')$, and then we can conclude as in the preceding case that for a suitable choice of t_2 the quotient t_2/t_1 is a non unit in $Q_{V'}(P')$. Using the results of section 5, especially d) and f), we conclude as follows:

If we put

$$t'_1 = t_1, \qquad t'_2 = t_2/t_1, \qquad t'_3 = t_3, \tag{10}$$

then t'_1, t'_2 and t'_3 are uniformizing parameters of $P'(V')$ and the quotient ring $Q_{V'}(P')$ consists of all quotients $\phi(t'_1, t'_2, t'_3)/\psi(t'_1, t'_2, t'_3)$, where ϕ and ψ are polynomials with coefficients in $Q_V(P)$ and where the "constant term" of ψ is a unit in $Q_V(P)$.

Here again, the equations (10) can be regarded locally (i.e. in regard to the pair of corresponding points P and P') as the equations of T. The form of these equations is the same as that of a monoidal transformation between two affine three-dimensional spaces (t_1, t_2, t_3) and (t'_1, t'_2, t'_3), over an algebraically closed field, the center being the line $t_1 = t_2 = 0$. The points P and P' play the role of the origins $(t) = 0$, $(t') = 0$.

3) We finally consider the case of a quadratic transformation T with center P, and we assume that $T[P]$ contains a point P' with the same residue field as P. By the same reasoning as in the preceding two cases and making use of the results of section 2, we conclude as follows:

For a suitable choice of the uniformizing parameters t_1, t_2, t_3 of $P(V)$, the elements

$$t_1' = t_1, \qquad t_2' = t_2/t_1, \qquad t_3' = t_3/t_1 \tag{11}$$

are uniformizing parameters of $P'(V')$. The quotient ring $Q_{V'}(P')$ consists of all quotients $\phi(t_1', t_2', t_3')/\psi(t_1', t_2', t_3')$, where ϕ and ψ are polynomials with coefficients in $Q_V(P)$ and where the "constant term" of ψ is a unit in $Q_V(P)$.

The form of the equations (11) is typical for a quadratic transformation between two three-dimensional affine spaces, over an algebraically closed field, the center of the transformation being the origin $t_1 = t_2 = t_3 = 0$.

8. Normal crossings

Let Δ be an irreducible curve on V and let P be a simple point of Δ. The prime ideal of Δ in $Q_V(P)$ has then a base consisting of two elements ω_1 and ω_2 whose leading forms are linear and linearly independent. That means that if t_1, t_2, t_3 are uniformizing parameters of $P(V)$, and if k^* denotes the residue field of P, then $\omega_i = f_i(t_1, t_2, t_3)$, where $f_i(z_1, z_2, z_3)$ is a linear form with coefficients in $Q_V(P)$ and where $f_1^*(z_1, z_2, z_3)$ and $f_2^*(z_1, z_2, z_3)$ are linearly independent over k^*. Here $f_i^*(z_1, z_2, z_3)$ denotes, as usual, the form in $k^*[z_1, z_2, z_3]$ whose coefficients are the P-residues of the coefficients of $f_i(z_1, z_2, z_3)$. The two forms f_1^*, f_2^* define a point in the projective plane $S_2^{k^*}$, or also a direction through the origin $z_1 = z_2 = z_3 = 0$ in an affine three-dimensional space over k^*. It is the *tangential direction of Δ at the point P.*

If P is a common point of two irreducible curves Δ and $\bar{\Delta}$ we shall say that P is a *normal crossing* of Δ and $\bar{\Delta}$ if P is simple point of both curves and if the tangential directions of Δ and $\bar{\Delta}$ at P are distinct.

If Δ and $\bar{\Delta}$ are two ν-fold curves on F we shall say that P is a normal crossing *on F* if P is a normal crossing and *if it is exactly ν-fold for F.*

LEMMA 8.1. *If P is a normal crossing of two irreducible curves Δ and $\bar{\Delta}$, then there exist uniformizing parameters t_1, t_2, t_3 of $P(V)$ such that (t_1, t_2) is a basis for the prime ideal of Δ in $Q_V(P)$ and (t_2, t_3) is a basis for the prime ideal of $\bar{\Delta}$ in $Q_V(P)$.*

PROOF. Let τ_1, τ_2 and τ_3, τ_4 be bases respectively for the ideals of Δ and $\bar{\Delta}$ in $Q_V(P)$. Since P is a normal crossing, three of the leading forms of the 4 elements τ_i must be linearly independent. Assuming that the leading forms of τ_1, τ_2, τ_3 are linearly independent, we will then have that τ_1, τ_2, τ_3 are uniformizing parameters of $P(V)$. We therefore can write $\tau_4 = a_1\tau_1 + a_2\tau_2 + a_3\tau_3$, $a_i \in Q_V(P)$. The elements a_1 and a_2 cannot be both zero at P, for otherwise the leading forms of τ_3 and τ_4 would be linearly dependent, contrary to the assumption that P is a simple point of $\bar{\Delta}$. Let, say, $a_2 \neq 0$ at P, and let us put $t_1 = \tau_1$, $t_2 = a_1\tau_1 + a_2\tau_2$, $t_3 = \tau_3$. Since a_2 is a unit in $Q_V(P)$, the elements

(t_1, t_2) form a basis for the prime ideal of Δ in $Q_V(P)$. On the other hand we have $t_2 = \tau_4 - a_3\tau_3$, and hence t_2 and t_3 form a basis for the prime ideal of $\bar{\Delta}$ in $Q_V(P)$, q.e.d.

Let now Δ and $\bar{\Delta}$ be two irreducible ν-fold curves of the surface F and let F' be the transform of F be the monoidal transformation of center Δ.

LEMMA 8.2. *If T carries Δ into a ν-fold curve Δ' of F' and if P is a normal crossing of Δ and $\bar{\Delta}$ on F, then the corresponding point P' of $\bar{\Delta}' = T[\bar{\Delta}]$ is a normal crossing of Δ' and $\bar{\Delta}'$ on F'.*

PROOF. We select the uniformizing parameters of $P(V)$ as in Lemma 8.1. By Theorem 2 of section 6 the transform $T[\Delta]$ is the irreducible curve Δ', and by b), section 7, there corresponds to P a unique point P' of Δ'. Since P' is ν-fold for F', the residue fields of P and P' coincide. We are therefore in position to apply the results of the preceding section. In order to decide which one of the two quotients t_1/t_2 and t_2/t_1 certainly belongs to $Q_{V'}(P')$, let us consider the principal ideal (ω) in $Q_V(P)$ which defines the surface F. Since both curves Δ and $\bar{\Delta}$ are ν-fold for F, ω must be expressable in the following form:

$$(12) \qquad \omega = \omega_0 t_2^{\nu} + \omega_1 t_2^{\nu-1} t_1 t_3 + \cdots + \omega_\nu (t_1 t_3)^\nu,$$

where the ω_i are in $Q_V(P)$ and where ω_0 *is not zero at* P, since $m_F(P) = \nu$. Thus the leading form of ω is $\overset{*}{\omega_0} z_2^{\nu}$, where $\overset{*}{\omega_0}$ is the P-residue of ω_0. Consequently [see f), section 5], $t_2/t_1 \in Q_{V'}(P')$, but t_1/t_2 is definitely not in $Q_{V'}(P')$. Therefore t_2/t_1 is a non unit in $Q_{V'}(P')$, and the equations (10) of the preceding section are valid.

The surface F' is defined in $Q_{V'}(P')$ by the principal ideal (ω'), where

$$(12') \qquad \omega' = \omega/t_1^{\nu} = \omega_0 t_2'^{\nu} + \omega_1 t_2'^{\nu-1} t_3' + \cdots + \omega_\nu t_3'^{\nu}.$$

The curve $\bar{\Delta}'$ is defined in $Q_{V'}(P')$ by the prime ideal (t_2', t_3') and is naturally ν-fold for F', as is also shown clearly by (12'). Therefore P' is a simple point of $\bar{\Delta}'$. As to Δ', we know that it lies on the surface Φ' $(= T_1[\Delta])$, and therefore [e), section 5], t_1' is one of the two generators of the prime ideal of Δ' in $Q_{V'}(P')$. Hence P' is a normal crossing of Δ' and $\bar{\Delta}'$ on F', as was asserted.

In the sequel we shall have occasion to follow up the monoidal transformation T_1, of center Δ, by a monoidal transformation T_1', of center $\bar{\Delta}'$. Now we could have first applied a monoidal transformation of center Δ' and then a monoidal transformation whose center is the transform of Δ. It turns out that interchanging the order in which the curves Δ and $\bar{\Delta}$ are used as centers of a monoidal transformation does not affect essentially the resulting composite transformation of the surface F, i.e. *to within a regular birational correspondence.* We proceed to state precisely the result.

We suppose then that Δ and $\bar{\Delta}$ are irreducible ν-fold curves of F and we apply to V a monoidal transformation T_1 of center Δ. We denote, as usual, by V' and F' the transforms of V and of F respectively, and by T—the transformation from F to F' induced by T_1. We suppose that $T[\Delta]$ is an irreducible ν-fold curve Δ' of F', and we denote the transform $T[\bar{\Delta}]$ by $\bar{\Delta}'$. We next apply to V'

a monoidal transformation $\overline{T}_1'$, of center $\overline{\Delta}'$, and we denote by V'' and F'' the transforms of V' and F' respectively, and by $\overline{T}'$—the transformation from F' to F'' which is induced by $\overline{T}_1'$. We suppose that $\overline{T}'[\overline{\Delta}']$ is irreducible and ν-fold for F''.

We now interchange the order in which the two curves Δ and $\overline{\Delta}$ have been used. We first apply to V a monoidal transformation $\overline{T}_1$ of center $\overline{\Delta}$, getting entities V^*, F^* and $\overline{T}$ whose meaning is similar to that of V', F' and T. Since T does not affect the quotient ring of $\overline{\Delta}$ we have $Q_F(\overline{\Delta}) = Q_{F'}(\overline{\Delta}')$, and since we have assumed that $\overline{T}'[\overline{\Delta}']$ is irreducible and ν-fold for F'', it follows therefore that also $\overline{T}[\overline{\Delta}]$ is irreducible and ν-fold for F^*. Let $\overline{T}[\overline{\Delta}]$ and $\overline{T}[\Delta]$ be denoted respectively by $\overline{\Delta}^*$ and Δ^*. We next apply to V^* a monoidal transformation T_1^* of center Δ^*, getting V^{**}, F^{**} and T^*. Since $\overline{T}$ does not affect the quotient ring of Δ, we have $Q_F(\Delta) = Q_{F^*}(\Delta^*)$, and since we have assumed that $T[\Delta]$ is irreducible and ν-fold for F', it follows that also $T^*[\Delta^*]$ is irreducible and ν-fold for F^{**}.

Lemma 8.3. *If P'' and P^{**} are two corresponding points of F'' and F^{**}, i.e. if P'' and P^{**} correspond to one and the same point P of F, then $Q_{F''}(P'') = Q_{F^{**}}(P^{**})$, except when P is a common point of Δ and $\overline{\Delta}$ which is not a normal crossing on F.*

Proof. Let P' and P^* be the points of F' and F^* respectively which correspond respectively to P'' and P^{**}. We shall first suppose that P is not a common point of Δ and $\overline{\Delta}$. If P does not lie on either curve, then all the transformations used above do not affect the quotient ring $Q_V(P)$ at all, and hence $Q_{V''}(P'') = Q_{V^{**}}(P^{**}) = Q_V(P)$. Assume now that $P \in \Delta$, $P \notin \overline{\Delta}$. Since $P \notin \overline{\Delta}$, we have $Q_F(P) = Q_{F^*}(P^*)$ and $P^* \in \Delta^*$. Therefore the monoidal transformations T and T^* are locally identical at P and P^* respectively, whence $Q_{F'}(P') = Q_{F^{**}}(P^{**})$. On the other hand, since $P \notin \overline{\Delta}$, we also have $P' \notin \overline{\Delta}'$, and therefore $Q_{F'}(P') = Q_{F''}(P'')$, and the assertion of the lemma is proved in this case.

Suppose now that P is a normal crossing, on F, of Δ and $\overline{\Delta}$. Locally, for the pair of corresponding points P and P', the equations of T_1 are given by (10), i.e.

$$t_1' = t_1, \quad t_2' = t_2/t_1, \quad t_3' = t_3. \tag{13'}$$

We know already that in $Q_{V'}(P')$ the curve $\overline{\Delta}'$ is given by the prime ideal (t_2', t_3'). From the form (12′) of the element ω' and from the fact that $\omega_0 \neq 0$ at P' (since $\omega_0 \neq 0$ at P), we conclude that $t_2'/t_3' \in Q_{V''}(P'')$. Since the residue fields of P and P'' are the same, we can find an element a in $Q_V(P)$ such that $t_2'/t_3' - a$ be a non unit in $Q_{V''}(P'')$. Now note that $(t_1, t_2) = (t_1, t_2 - at_1t_3)$ and $(t_2, t_3) = (t_2 - at_1t_3, t_3)$, all ideals being in $Q_V(P)$. Hence we may replace t_2 by $t_2 - at_1t_3$. With this choice of t_2, t_2'/t_3' will be a non unit in $Q_{V''}(P'')$ since $t_2' - at_3' = (t_2 - at_1t_3)/t_1$. Hence the equations of $\overline{T}_1'$ are as follows:

$$t_1'' = t_1', \quad t_2'' = t_2'/t_3', \quad t_3'' = t_3', \tag{13''}$$

where t_1'', t_2'', t_3'' are uniformizing parameters of $P''(F'')$.

In a similar fashion we find that the equations of $\overline{T}_1$ are

$$(13^*) \qquad t_1^* = t_1, \quad t_2^* = t_2/t_3, \quad t_3^* = t_3,$$

where t_1^*, t_2^*, t_3^* are the uniformizing parameters of $P^*(V^*)$, and that the equations of T_1^* are as follows:

$$(13^{**}) \qquad t_1^{**} = t_1^*, \quad t_2^{**} = t_2^*/t_1^*, \quad t_3^{**} = t_3^*,$$

where t_1^{**}, t_2^{**}, t_3^{**} are uniformizing parameters of $P^{**}(V^{**})$. From (13′), (13″), (13*) and (13**) we conclude that

$$(14) \qquad t_1'' = t_1^{**} = t_1, \quad t_2'' = t_2^{**} = t_2/t_1t_3, \quad t_3'' = t_3^{**} = t_3,$$

and this shows the complete identity of the composite transformations $T_1\overline{T}_1'$ and $\overline{T}_1T_1^*$ locally, at (P, P'') and (P, P^{**}). This completes the proof of the Lemma.

Lemma 8.4. *Let Δ be an irreducible ν-fold curve of F and let P be a simple point of Δ which is exactly ν-fold for F. If a quadratic transformation T, of center P, transforms P into a ν-fold curve Γ' of F', then the transform Δ' of Δ intersects Γ' in one point only, and that point is a normal crossing of Δ' and Γ' on F'.*

Proof. We know from b), section 2, that the points of Φ' are in (1, 1) correspondence with the directions about P (since each direction about P is represented by a point of the projective plane S). Since P is simple for Δ, it follows that Δ' intersects Φ', and hence also Γ', in only one point. Let P' be the common point of Δ' and Γ'. If t_1, t_2, t_3 are uniformizing parameters of $P(V)$ such that t_2, t_3 are uniformizing parameters of $\Delta(V)$, then the equations of T_1 are given by (11), section 7, where t_1', t_2', t_3' are uniformizing parameters of $P'(V')$. The curve Δ' is given in $Q_{V'}(P')$ by the ideal (t_2', t_3'), while the surface Φ' is defined by the ideal (t_1'). Since $\Gamma' \subset \Phi'$ and Γ' has a simple point at P' (Theorem 1, section 3), and since moreover P' is exactly ν-fold for F', it follows that P' is a normal crossing of Δ' and Γ', as asserted.

Part II

Local Reduction Theorem for Surfaces Embedded in a V_3

9. Reduction of the singularities of an algebraic curve by quadratic transformations

In the sequel we shall have to make use of the theorem that the singularities of any higher space curve can be resolved by quadratic transformations of the ambient space. This theorem is well-known in the classical case. In this section we shall prove this theorem for arbitrary ground fields. It is clear that all depends on proving the following theorem:

Theorem 4. *If*

$$\Gamma_1, \Gamma_2, \cdots, \Gamma_i, \cdots$$

is an infinite sequence of irreducible algebraic curves (over a common ground field k) and if $P_1, P_2, \cdots, P_i, \cdots$ is a corresponding sequence of points such that: 1) P_i is a point of Γ_i; 2) Γ_{i+1} is obtained from Γ_i by a quadratic transformation T_i of center P_i; 3) the point P_{i+1} is one of the points of Γ_{i+1} which correspond to P_i

(there is only a finite number of such points), then if i is sufficiently high, P_i is a simple point of Γ_i.

Proof. Let $\mathfrak{Z}_i$ be the quotient ring $Q_{\Gamma_i}(P_i)$, so that we have

$$\mathfrak{Z}_1 \subseteqq \mathfrak{Z}_2 \subseteqq \cdots \subseteqq \mathfrak{Z}_i \subseteqq \cdots,$$

and let Ω be the union of the rings $\mathfrak{Z}_i$. Any non unit in $\mathfrak{Z}_1$ is also a non unit in $\mathfrak{Z}_2, \mathfrak{Z}_3, \cdots$, and consequently Ω is a proper ring (i.e., not a field). Therefore Ω is contained in the valuation ring R_v of at least one valuation v of the field of algebraic functions of which the curves Γ_i are projective models.

Lemma 9.1. *If $\Omega \subseteqq R_v$, then $\Omega = R_v$.*

Proof. Let $\mathfrak{P}$ be the ideal of non units of R_v and let $\mathfrak{m}_i$ be the ideal of non units of $\mathfrak{Z}_i$, so that $\mathfrak{P} \cap \mathfrak{Z}_i = \mathfrak{m}_i$.

Let ξ_1/η_1 be an arbitrary element of R_v, where $\xi_1, \eta_1 \epsilon \mathfrak{Z}_1$. If $\eta_1 \not\equiv 0(\mathfrak{m}_1)$, then $\xi_1/\eta_1 \epsilon \mathfrak{Z}_1$. Suppose that $\eta_1 \equiv 0(\mathfrak{m}_1)$. Then necessarily also $\xi_1 \equiv 0(\mathfrak{m}_1)$. Considerations similar to those developed in section 2 show immediately that the extended ideal $\mathfrak{Z}_2 \cdot \mathfrak{m}_1$ is a principal ideal, say $\mathfrak{Z}_2 \cdot \mathfrak{m}_1 = \mathfrak{Z}_2 \cdot \beta_2$, $\beta_2 \epsilon \mathfrak{m}_2$ (see also [5], p. 679, formula (33)). Let $\xi_1 = \xi_2\beta_2$, $\eta_1 = \eta_2\beta_2$, where $\xi_2, \eta_2 \epsilon \mathfrak{Z}_2$, whence $\xi_1/\eta_1 = \xi_2/\eta_2$. If $\eta_2 \not\equiv 0(\mathfrak{m}_2)$, then $\xi_2/\eta_2 \epsilon \mathfrak{Z}_2$. In the contrary case we will have by a similar argument: $\xi_2 = \xi_3\beta_3$, $\eta_2 = \eta_3\beta_3$, $\beta_3 \equiv 0(\mathfrak{m}_3)$. Since $\beta_2 = 0(\mathfrak{m}_2)$, we have $v(\beta_2) > 0$, whence $v(\xi_1) > v(\xi_2) \geqq 0$ and $v(\eta_1) > v(\eta_2) \geqq 0$. Similarly, $v(\xi_2) > v(\xi_3) \geqq 0$ and $v(\eta_2) > v(\eta_3) \geqq 0$. Since the value group of v is the additive group of integers, it follows that if i is sufficiently high then $v(\eta_i) = 0$, and then we will have $\eta_i \not\equiv 0(\mathfrak{m}_i)$, $\xi_1/\eta_1 = \xi_i/\eta_i \epsilon \mathfrak{Z}_i$. Thus every element of R_v is contained in one of the quotient rings $\mathfrak{Z}_i$, and this completes the proof of the lemma.

Each point P_i is the center of only a finite number of valuations (P_i is the origin of only a finite number of branches of Γ_i). From the preceding lemma we therefore may conclude as follows:

a) *If i is sufficiently high then R_v is the only valuation ring containing $\mathfrak{Z}_i$* (i.e. P_i is the origin of only one branch of Γ_i).

The residue field $R_v/\mathfrak{P}$ of the valuation v is a finite extension of the ground field k, and on the other hand this residue field must be, by Lemma 9.1, the union of the residue fields $\mathfrak{Z}_i/\mathfrak{m}_i$. Since these form an ascending sequence of fields, it follows that:

b) *If i is sufficiently high then $\mathfrak{Z}_i/\mathfrak{m}_i = \mathfrak{Z}_{i+1}/\mathfrak{m}_{i+1} = \cdots =$ residue field of v.*

From Lemma 9.1 it also follows that

c) *If i is sufficiently high then $\mathfrak{m}_i$ contains elements of value 1 in v; in other words: $R_v \cdot \mathfrak{m}_i = \mathfrak{P}$, for i sufficiently high.*

In view of a), b) and c), Theorem 4 is a consequence of the following lemma:

Lemma 9.2. *A point P of an irreducible algebraic curve Γ is a simple point of Γ if the following conditions are satisfied:*[5]

[5] It is obvious that the conditions a), b) and c) are also necessary, since a point of an algebraic curve is a simple point of the curve if and only if the quotient ring of the point is a valuation ring.

a) *The point P is the center of only one valuation of the field of rational functions on Γ.*

b) *If v is the valuation of center P, then the residue field of P coincides with the residue field of the valuation v.*

c) *The quotient ring $Q_\Gamma(P)$ contains elements of value 1.*

PROOF. Let $\mathfrak{J} = Q_\Gamma(P)$ and let $\bar{\mathfrak{J}}$ denote the integral closure of $\mathfrak{J}$ in its quotient field. Since $\bar{\mathfrak{J}}$ is the intersection of valuation rings, it follows by a) that $\bar{\mathfrak{J}} = R_v$. Since $\mathfrak{J}$ is the quotient ring of a prime ideal of a finite integral domain (namely of the ring generated over k by the non-homogeneous coördinates of the general point of Γ), it is known that $\bar{\mathfrak{J}}$ is a finite $\mathfrak{J}$-module (see F. K. Schmidt [3]). Hence the conductor $\mathfrak{C}$ of $\bar{\mathfrak{J}}$ with respect to $\mathfrak{J}$ is not the zero ideal. The conductor $\mathfrak{C}$ is therefore a primary ideal in $\mathfrak{J}$, whose associated prime ideal is the ideal $\mathfrak{m}$ of non-units in $\mathfrak{J}$; unless $\mathfrak{C} = \mathfrak{J}$, in which case $\mathfrak{J} = \bar{\mathfrak{J}} = R_v$, and P is then indeed a simple point of Γ. In either case, since $\mathfrak{J}$ is a chain-theorem ring, we will have $\mathfrak{m}^\rho \equiv 0(\mathfrak{C})$, for some integer ρ. Since $\mathfrak{C}$ is also an ideal in $\bar{\mathfrak{J}}$, the congruence $\mathfrak{m}^\rho \equiv 0(\mathfrak{C})$ implies, by c), that *every element of R_v whose value is $\geqq \rho$ is necessarily an element of $\mathfrak{J}$.* Let now t be an element of $\mathfrak{J}$ such that $v(t) = 1$, in view of c), and let ξ be an arbitrary non unit in $\mathfrak{J}$, whence $v(\xi) = m > 0$. By b), there exists an element α in $\mathfrak{J}$ whose P-residue coincides with the residue of ξ/t^m in the valuation v. Therefore, if we put

$$\xi = \alpha t^m + \xi_1,$$

then we will have $v(\xi_1) = m_1 > m$. In a similar fashion we can find an element α_1 in $\mathfrak{J}$ such that $v(\xi_1 - \alpha_1 t^{m_1}) = m_2 > m_1$, whence

$$\xi = \alpha t^m + \alpha_1 t^{m_1} + \xi_2, \qquad m < m_1 < m_2 = v(\xi_2).$$

Let

$$\xi = \alpha t^m + \alpha_1 t^{m_1} + \alpha_2 t^{m_2} + \cdots + \alpha_h t^{m_h} + \xi_{h+1}, \tag{15}$$

where

$$1 \leqq m < m_1 < m_2 < \cdots < m_h < m_{h+1} = v(\xi_{h+1})$$

and $\alpha, \alpha_1, \alpha_2, \cdots, \alpha_h, \xi_{h+1} \in \mathfrak{J}$. Let us take h so high that $m_{h+1} \geqq \rho + 2$. Then $v(\xi_{h+1}/t^2) \geqq \rho$, whence $\xi_{h+1}/t^2 \in \mathfrak{J}$ and consequently $\xi_{h+1} \equiv 0(\mathfrak{J}\cdot t^2)$. From (15) we therefore conclude that $\xi \equiv 0(\mathfrak{J}\cdot t)$. Since this holds for every element of $\mathfrak{m}$, it follows that $\mathfrak{m} = \mathfrak{J}\cdot t$, whence P is a simple point, with t as uniformizing parameter, q.e.d.

Let Γ be an irreducible ν-fold curve of a surface F, $\nu > 1$, and let $F^{(1)}$ be a transform of F by a monoidal transformation of center Γ. To Γ there will correspond on $F^{(1)}$ a finite number of irreducible curves. Suppose that one of these curves, say $\Gamma^{(1)}$, is still ν-fold for $F^{(1)}$. We apply to $F^{(1)}$ a monoidal transformation of center $\Gamma^{(1)}$, getting a surface $F^{(2)}$. If $F^{(2)}$ still contains a ν-fold curve $\Gamma^{(2)}$ which corresponds to $\Gamma^{(1)}$, we repeat the process. Suppose that we get in this fashion a sequence of consecutive ν-fold curves

$$\Gamma, \Gamma^{(1)}, \Gamma^{(2)}, \cdots, \Gamma^{(i)}, \cdots$$

all ν-fold for their respective carriers $F, F^{(1)}, F^{(2)}, \cdots, F^{(i)}, \cdots$, where $F^{(i+1)}$ is obtained from $F^{(i)}$ by a monoidal transformation of center $\Gamma^{(i)}$ and where $\Gamma^{(i)}$ corresponds to $\Gamma^{(i+1)}$ in that monoidal transformation.

THEOREM 4′. *The sequence* $\Gamma, \Gamma^{(1)}, \Gamma^{(2)}, \cdots$ *is necessarily finite.*

PROOF. Let ξ be an element of the field of rational functions on F such that the Γ-residue of ξ is a transcendental with respect to the ground field k. Over the extended ground field $k(\xi)$ the surface F becomes a curve, Γ becomes a point, and our successive *monoidal* transformations become quadratic. It is thus seen that Theorem 4′ is nothing but Theorem 4 in a different "dimension terminology."

10. A local reduction theorem for algebraic surfaces

Let F be an irreducible algebraic surface embedded in a V_3, and let v be a zero-dimensional valuation of the field of rational functions on F such that the center of v on F is a point P which is simple for V_3. Let P be a ν-fold point of F. We make the following assumption: *If P is not an isolated ν-fold point of F then it is either a simple point or a normal crossing of the total (possibly reducible) ν-fold curve of F.* By a *permissible transformation* of V_3 we shall mean either a quadratic transformation of center P (regardless whether P is or is not an isolated ν-fold point) or a monoidal transformation whose center is a ν-fold curve through P.

Let T_1 be a permissible transformation of V_3 and let V_3' and F' be the transforms of V_3 and F respectively, under T_1. Let T be the birational transformation between F and F' induced by T_1. *We suppose that the center P' of v on F' is still ν-fold* and we consider various possible cases.

a) *P is an isolated ν-fold point of F.* In this case T_1 is necessarily quadratic and any ν-fold curve through P' can only be a new ν-fold curve created by the transformation T, i.e. it must belong to $T[P]$. It then follows from Theorem 1, section 3, that if P' is not an isolated ν-fold point it is a simple point of the ν-fold curve of F'.

b) *P is a simple point of the ν-fold curve of F.* Let Δ be the irreducible component of the ν-fold curve which passes through P. If T_1 is quadratic, then only the following cases are possible: 1) P' is an isolated ν-fold point P'; 2) $T[\Delta]$ is the only irreducible ν-fold curve of F' which passes through P'; P' is a simple point of $T[\Delta]$ since P is simple for Δ; 3) $T[P]$ is ν-fold for F' and is the only irreducible ν-fold curve of F' passing through P'; 4) $T[P]$ is ν-fold for F', and P' belongs to both curves $T[P]$ and $T[\Delta]$; in this case P' is a normal crossing, by Lemma 8.4.

Now suppose that T_1 is a monoidal transformation of center Δ. If $T[\Delta]$ is not ν-fold for F', then P' is an isolated ν-fold point. If $T[\Delta]$ is ν-fold, then it is irreducible (Theorem 2, section 6), and P' must be a simple point of $T[\Delta]$ (Corollary to Theorem 2), and there are no other ν-fold curves through P'.

c) *P is a normal crossing of two ν-fold curves Δ and $\bar{\Delta}$.* If T_1 is a quadratic transformation, then P' cannot belong to both curves $T[\Delta]$ and $T[\bar{\Delta}]$ since Δ

and $\bar{\Delta}$ have at P distinct tangent lines. This case does not differ from case b) considered above. If T_1 is a monoidal transformation of center Δ, then in addition to the possibilities which have arisen in the preceding case b) we have also the possibility that P' is a normal crossing of the two ν-fold curves $T[\Delta]$ and $T[\bar{\Delta}]$ (see Lemma 8.2).

We conclude that in all cases the center P' of the valuation v on F' is again either an isolated ν-fold point, or a simple point of the ν-fold curve of F' or a normal crossing of the ν-fold curve of F'. All this provided that P satisfy these same conditions, that P' be a ν-fold point and that T_1 be a permissible transformation.

We next apply to V_3' and F' a permissible transformation T_1', relative to the point P', getting V_3'', F'' and P''. If P'' is still ν-fold for F'', we apply a permissible transformation T_1'' to V_3'', relative to the point P'', and so we continue indefinitely as long as the center of the valuation v remains a ν-fold point. In this fashion we get a sequence of surfaces $F, F', F'', \cdots, F^{(i)}, \cdots$, where $F^{(i+1)}$ is obtained from $F^{(i)}$ by a permissible transformation $T_1^{(i)}$ relative to the center $P^{(i)}$ of v on $F^{(i)}$ where it is supposed that $P^{(i)}$ is still ν-fold for $F^{(i)}$. We observe that whenever $P^{(i)}$ is not an isolated ν-fold point of its carrier $F^{(i)}$, then, in view of our definition of a permissible transformation, $T_1^{(i)}$ may be either quadratic or monoidal. We have therefore, in general, a certain degree of freedom in building up the sequence of successive transforms $F^{(i)}$.

The main purpose of this second part of the paper is to establish the following *local reduction theorem*:

THEOREM 5. *By a suitably chosen sequence of permissible transformations it is possible to obtain a birational transform F^* of F on which the center of the valuation v is a point of multiplicity less than ν.*

This theorem naturally implies the theorem of uniformization of valuations which we have first proved in [5] for surfaces over an algebraically closed ground field of characteristic zero and later on, in [6], for higher varieties over arbitrary ground fields of characteristic zero. However, our present local reduction theorem is stronger than the pure existence theorem ("there exists a birational transform of F on which the center of the given valuation is a simple point") proved in the quoted papers, since it calls for a proof that the valuation can be uniformized by birational transformations of a *specified type*. Another difference is this: in the proof of the general uniformization theorem for surfaces, it was sufficient to carry out explicitly the uniformization for surfaces embedded in a linear 3-space. In the present case our surface is embedded in an arbitrary V_3 and it is essential that the uniformization be carried out by birational transformations of the ambient V_3.

The broad lines of the proof of Theorem 5 are essentially the same as those of our old proof of the uniformization of valuations. However, their adaptation to the new set-up calls for a technique in which the local character of the problem comes to the fore, for the controlling ring is now not a ring of polynomials in three variables but the quotient ring of a simple point P of V_3. The resulting re-

organization of the proof actually implies an improvement on our old proof, especially when the ground field k is not algebraically closed.

In the following sections we take up the various types of zero-dimensional valuations and prove Theorem 5 for each type separately. As was pointed out in the Introduction, our proof of Theorem 5 for valuations of rank 1 applies only to fields of characteristic zero. For valuations of rank 2, our proof applies to arbitrary fields.

11. Some preliminary lemmas

Let V be an irreducible algebraic variety and let v be a zero-dimensional valuation of the field of rational functions on V. The valuation v determines an infinite sequence of successive birational transforms of V, say

$$V, V_1, V_2, \cdots, V_i, \cdots$$

where V_{i+1} is obtained by applying to V_i a quadratic transformation T_i whose center is the center P_i of v on V_i. Let Γ be an irreducible algebraic curve on V passing through P and let $T[\Gamma] = \Gamma_1$, $T_1[\Gamma_1] = \Gamma_2, \cdots, T_i[\Gamma_i] = \Gamma_{i+1}, \cdots$.

LEMMA 11.1. *If $P_i \in \Gamma_i$ for all i, then v is compounded with a valuation whose center on V is the curve Γ.*

PROOF. Let $\mathfrak{J}_i = Q_{V_i}(P_i)$, $\bar{\mathfrak{J}}_i = Q_{\Gamma_i}(P_i)$ and let $\mathfrak{m}_i$ and $\bar{\mathfrak{m}}_i$ be respectively the ideal of non units in $\mathfrak{J}_i$ and $\bar{\mathfrak{J}}_i$. The ideal $\mathfrak{J}_{i+1}\cdot\mathfrak{m}_i$ is principal [cf. section 2; also [5], p. 679, formula (33)], say $\mathfrak{J}_{i+1}\cdot\mathfrak{m}_i = \mathfrak{J}_{i+1}\cdot\beta_{i+1}$. If $\bar{\beta}_{i+1}$ is the Γ_{i+1}-residue of β_{i+1}, then $\bar{\beta}_{i+1} \neq 0$ and $\bar{\mathfrak{J}}_{i+1}\cdot\bar{\mathfrak{m}}_i = \bar{\mathfrak{J}}_{i+1}\cdot\bar{\beta}_{i+1}$. Let now ξ and η be elements of $\mathfrak{J}$ and let us assume that ξ is zero on Γ and $\eta \neq 0$ on Γ. We have then $\xi = \xi_1\beta_1$, $\xi_1 \in \mathfrak{J}_1$, $\xi_1 = 0$ on Γ_1. If η is a non unit in $\mathfrak{J}$ we can also write $\eta = \eta_1\beta_1$, $\eta_1 \in \mathfrak{J}_1$. If η_1 is a non unit in $\mathfrak{J}_1$, we define the element η_2 in $\mathfrak{J}_2$ by the relation $\eta_1 = \eta_2\beta_2$, and we also put $\xi_1 = \xi_2\beta_2$. More generally, if the elements $\eta_1, \eta_2, \cdots, \eta_s$ are definable by the above construction, and if η_s is a non unit in $\mathfrak{J}_s$, we put $\xi_s = \xi_{s+1}\beta_{s+1}$, $\eta_s = \eta_{s+1}\beta_{s+1}$. If $\bar{\xi}_i$, $\bar{\eta}_i$ denote the Γ_i-residues of ξ_i and η_i respectively, then $\bar{\xi}_i = 0$ and $\bar{\eta}_i = \bar{\eta}_{i+1}\bar{\beta}_{i+1}$ $(i = 1, 2, \cdots, s)$. Since $\bar{\mathfrak{J}}_{i+1}\cdot\bar{\mathfrak{m}}_i = \bar{\mathfrak{J}}_{i+1}\cdot\bar{\beta}_{i+1}$, the proof of Lemma 9.1 shows that for some integers s we must get an element $\bar{\eta}_{s+1}$ which is a unit in $\bar{\mathfrak{J}}_{s+1}$. Then η_{s+1} is a unit in $\mathfrak{J}_{s+1}$, and consequently $v(\xi_{s+1}) > v(\eta_{s+1})$, since ξ_{s+1} is zero at P_{s-1}. Since $\xi_{s+1}/\eta_{s+1} = \xi/\eta$, we conclude that $v(\xi) > v(\eta)$. We have thus proved the following: *if $\xi, \eta \in \mathfrak{J}$, $\xi = 0$ on Γ and $\eta \neq 0$ on Γ, then $v(\xi) > v(\eta)$.* This shows that *the valuation v is of rank greater than* 1, for if η is such that $\eta = 0$ at P, $\eta \neq 0$ on Γ then $v(\eta) > 0$ and $v(\xi) > v(\eta^n)$, for any positive integer n.

Let m be the rank of v and let

$$\mathfrak{P} \supset \mathfrak{P}_1 \supset \cdots \supset \mathfrak{P}_{m-1} \neq (0)$$

be the strictly descending chain of the prime ideal of the valuation ring R_v, where $\mathfrak{P}$ is the ideal of non units in R_v. Since P is the center of v, we must have: $\mathfrak{P} \cap \mathfrak{J} = \mathfrak{m}$. On the other hand, $\mathfrak{P}_{m-1} \cap \mathfrak{J} \neq \mathfrak{m}$, since, as we have just seen, there exist pairs of elements ξ and η in $\mathfrak{m}$ such that $v(\xi)$ is greater than any

positive multiple of $v(\eta)$; for such a pair of elements it is true that η cannot belong to $\mathfrak{P}_{m-1}$. Let therefore s be an integer, $1 \leqq s \leqq m - 1$, with the property:

$$\mathfrak{P}_{s-1} \cap \mathfrak{J} = \mathfrak{m}, \qquad \mathfrak{P}_s \cap \mathfrak{J} = \mathfrak{p} \neq \mathfrak{m}.$$

From the preceding considerations it follows that every element ξ of $\mathfrak{J}$ which vanishes on Γ must belong to $\mathfrak{P}_s$, hence also to $\mathfrak{p}$. Hence the prime ideal $\mathfrak{p}$ is at most one-dimensional, and since $\mathfrak{p} \neq \mathfrak{m}$ we conclude that $\mathfrak{p}$ *is the prime ideal of the curve* Γ. If we now consider the valuation v_1 of the field of rational functions on V whose valuation ring is the quotient ring of $\mathfrak{P}_s$ in R_v, then v is compounded with v_1 and from the fact that $\mathfrak{p}$ is the prime ideal of Γ it follows that Γ is the center of v_1, q.e.d.

The following is an application of the preceding lemma. Let us suppose that P is a simple point of V, whence also P_i is a simple point of V_i. Let $t_1, t_2, \cdots, t_r$ be uniformizing parameters of $P(V)$. The point P_1 corresponds to a definite tangential direction of V at P, i.e. to a definite point $\bar{P}$ of a projective $(r - 1)$-dimensional space S over the residue field of P (cf. section 8 for the case $r = 3$; the general case is treated exactly in the same fashion). If $z_1, z_2, \cdots, z_r$ are homogeneous coordinates in S and if, say, $z_1 \neq 0$ at $\bar{P}$, then $\mathfrak{J}_1 \cdot \mathfrak{m} = \mathfrak{J}_1 \cdot t_1$ [cf. section 2, d)]. Let ω be an element of $\mathfrak{J}$ and let us suppose that ω is exactly divisible by $\mathfrak{m}^\nu$. Then ω is divisible by t_1^ν in $\mathfrak{J}_1$, say $\omega = t_1^\nu \cdot \omega_1$. We call ω_1 the *transform of* ω *by the quadratic transformation* T, or briefly: *the* T-*transform of* ω.

If $\nu \geqq 1$ then the ideal $\mathfrak{J} \cdot \omega$ defines a pure $(r - 1)$-dimensional subvariety W of V which passes through P (W may be reducible and may possess multiple components). The ideal $\mathfrak{J}_1 \cdot \omega_1$ defines that part of the variety $T[W]$ which passes through P_1 (see Lemma 3.1).

We now suppose that V is a surface ($r = 2$). Let $\omega_1, \omega_2, \cdots, \omega_i, \cdots$ be the successive transforms of ω by $T, T_1, \cdots$. Let $\Gamma^{(1)}, \Gamma^{(2)}, \cdots, \Gamma^{(h)}$ be the irreducible components of the algebraic curve defined on V by the ideal $\mathfrak{J} \cdot \omega$. Suppose that each ω_i is a non unit in the corresponding ring $\mathfrak{J}_i$, i.e. $\omega_i = 0$ at P_i. Then for at least one of the curves $\Gamma^{(i)}$, say for $\Gamma^{(1)}$, it must be true that $T_i[\Gamma^{(1)}]$ passes through P_i. But then, by Lemma 11.1, we conclude that the valuation v must be of rank 2 and must be compounded with a divisor whose center is the curve $\Gamma^{(1)}$. We therefore can state the following

LEMMA 11.2. *Let v be a zero-dimensional valuation of the field of rational functions on a given surface F, and let*

$$F_1, F_2, \cdots, F_i, \cdots$$

be the sequence of quadratic transforms of F determined by the valuation v [i.e. let F_{i+1} $(i \geqq 0)$ be the transform of F_i by a quadratic transformation whose center P_i on F_i is the center of v]. We assume that the center P of v on F is a simple point of F. Under the hypothesis that v is of rank 1, it is true that if ω is any element of $Q_F(P)$ and if $\omega_1, \omega_2, \cdots, \omega_i, \cdots$ are the successive transforms of ω by $T, T_1, T_2, \cdots$ (i.e., if ω_{i+1} is the T_i transform of ω_i), then for i sufficiently high the element ω_i is a unit in $Q_{F_i}(P_i)$.

In the next lemma, r is again arbitrary, although we shall actually use the lemma only for $r = 2, 3$.

LEMMA 11.3. *If $\omega \in Q_V(P)$, where P is a simple point of V, and if $\omega_1, \omega_2, \cdots$ are the successive quadratic transforms of ω by $T, T_1, T_2, \cdots$, then*

$$\omega = \epsilon_i t_{1i}^{\rho_{1i}} t_{2i}^{\rho_{2i}} \cdots t_{ri}^{\rho_{ri}} \omega_i, \tag{16}$$

where ϵ_i is a unit in $Q_{V_i}(P_i)$ and where $t_{1i}, t_{2i}, \cdots, t_{ri}$ form a set of uniformizing parameters of $P_i(V_i)$, while the exponents ρ_{ji} are non-negative rational integers.

PROOF. The lemma is true for $i = 1$, since if $\omega = t_1^{\nu}\omega_1$ then t_1 is one of a set of r uniformizing parameters of $P_1(V_1)$ [see section 2, e)]. We assume that the lemma is true for $i = n$ and we prove it for $i = n + 1$. Without loss of generality we may assume that

$$\omega_i = t_{1i}^{\nu_i}\omega_{i+1}, \tag{17}$$

and that t_{1i} is one of a set of r uniformizing parameters of $P_{i+1}(V_{i+1})$. We know from section 2 that the quotients t_{ji}/t_{1i}, $j = 2, 3, \cdots, r$, belong to $Q_{V_{i+1}}(P_{i+1})$. Without loss of generality we may assume that the first $s - 1$ of these quotients ($1 \leqq s \leqq r$) are non units, while the remaining quotients are units in $Q_{V_{i+1}}(P_{i+1})$. From the considerations developed in section 2 it follows immediately that the elements $t_{1i}, t_{2i}/t_{1i}, \cdots, t_{si}/t_{1i}$ form a subset of a set of uniformizing parameters of $P_{i+1}(V_{i+1})$. Hence we may put

$$t_{1,i+1} = t_{1i}, \qquad t_{2,i+1} = t_{2i}/t_{1i}, \cdots, t_{s,i+1} = t_{si}/t_{1i},$$

and we get from (16) and (17)

$$\omega = \epsilon_{i+1} t_{1,i+1}^{\rho_{1,i+1}} t_{2,i+1}^{\rho_{2,i+1}} \cdots t_{s,i+1}^{\rho_{s,i+1}} \omega_{i+1},$$

where

$$\rho_{1,i+1} = \rho_{1i} + \rho_{2i} + \cdots + \rho_{ri} + \nu_i,$$

$$\rho_{j,i+1} = \rho_{ji}, \qquad j = 2, 3, \cdots, s,$$

and where

$$\epsilon_{i+1} = \epsilon_i \prod_{j=s+1}^{r} (t_{ji}/t_{1i})^{\rho_{ji}},$$

whence ϵ_{i+1} is a unit in $Q_{V_{i+1}}(P_{i+1})$, q.e.d.

The next—and last—lemma of this section refers to a sequence of permissible birational transformations *all elements of which are quadratic transformations* (see section 10). Using the notations of section 10, let $\mathfrak{J} = Q_V(P)$, $\mathfrak{J}_i = Q_{V^{(i)}}(P^{(i)})$, and let the surface F be defined in $\mathfrak{J}$ to be the prime principal ideal $\mathfrak{J}\cdot\omega$.

LEMMA 11.4. *If the points $P, P^{(1)}, P^{(2)}, \cdots$ are all ν-fold for their respective carriers $F, F^{(1)}, F^{(2)}, \cdots$ (all the successive transformations $T, T^{(1)}, T^{(2)}, \cdots$ being quadratic) and if the leading form of ω is not a ν^{th} power (of a linear form), then the*

valuation v is discrete (and is therefore defined by an algebraic or analytical branch through P).

PROOF. Let t_1, t_2, t_3 be uniformizing parameters of $P(V)$ and let $F(z_1, z_2, z_3)$ be the leading form of ω; here F is a form of degree ν whose coefficients are in the residue field K of P. Since $P^{(1)}$ is a ν-fold point of $F^{(1)}$, and since, by hypothesis, F is not the ν^{th} power of a linear form, it follows from Lemma 7.1 and Theorem 3 (section 7), that F is a binary form of two linear forms in $K[z_1, z_2, z_3]$. Without loss of generality we may assume that $F = G(z_2 - a_1 z_1, z_3 - b_1 z_1)$, where where $a_1, b_1 \in K$ and G is a form with coefficients in K. The plane curve $F(z_1, z_2, z_3) = 0$ consists of straight lines through the point $z_1:z_2:z_3 = 1:a_1:b_1$, and since this curve is, by hypothesis, not a ν-fold line, the above point is the only ν-fold point of the curve and therefore represents the tangential direction of V at P which corresponds to the ν-fold point $P^{(1)}$. It follows that the local equations of the quadratic transformation T are as follows; [cf. equations (11), section 7]:

$$(18) \qquad t_1^{(1)} = t_1, \qquad t_2^{(1)} = \frac{t_2 - \alpha_1 t_1}{t_1} \qquad t_3^{(1)} = \frac{t_3 - \beta_1 t_1}{t_1},$$

where α_1 and β_1 are any two elements of $Q_V(P)$ whose P-residues are respectively a_1 and b_1 and where $t_1^{(1)}, t_2^{(1)}, t_3^{(1)}$ are uniformizing parameters of $P^{(1)}(V^{(1)})$. Let $g(u_1, u_2)$ be a form with coefficients in $\mathfrak{J}$ such that the reduced form mod P is $G(u_1, u_2)$. Then it is clear that ω and $g(t_2 - \alpha_1 t_1, t_3 - \beta_1 t_1)$ have the same leading form. Consequently

$$\omega = g(t_2 - \alpha_1 t_1, t_3 - \beta_1 t_1) + \zeta, \qquad \zeta \equiv 0(\text{mod } \mathfrak{m}^{\nu+1}),$$

where $\mathfrak{m}$ is the ideal of non units in $\mathfrak{J}$.

The surface $F^{(1)}$ is defined by the principal ideal $\mathfrak{J}_1 \cdot \omega_1$, where ω_1 is the T-transform of ω. Using (18) we find

$$(19) \qquad \omega_1 = g(t_2^{(1)}, t_3^{(1)}) + t_1^{(1)} \zeta_1, \qquad \zeta_1 \in \mathfrak{J}_1.$$

Let $F_1(z_1^{(1)}, z_2^{(1)}, z_3^{(1)})$ denote the leading form of ω_1. By hypothesis, F_1 is of degree ν, since $P^{(1)}$ is ν-fold for $F^{(1)}$. Hence it follows from (19) that

$$(20) \qquad F_1(0, z_2^{(1)}, z_3^{(1)}) = G(z_2^{(1)}, z_3^{(1)}),$$

whence $F_1(z_1^{(1)}, z_2^{(1)}, z_3^{(1)})$ is not the ν^{th} power of a linear form. Therefore we have now at $P^{(1)}$ the same situation as at P. Since $P^{(2)}$ is also ν-fold for $F^{(2)}$, we conclude in view of (20) and recalling that the residue field of $P^{(1)}$ is the same as that of $P^{(1)}$ (Theorem 3, section 7), that $F_1(z_1^{(1)}, z_2^{(1)}, z_3^{(1)})$ must be a binary form of two linear forms in the $z_i^{(1)}$, all the coefficients being in K. In view of (20) these two linear forms can be assumed to be $z_2^{(1)} - a_2 z_1^{(1)}$, $z_3^{(1)} - b_2 z_1^{(1)}$, $a_2, b_2 \in K$. If α_2, β_2 are elements of $\mathfrak{J}$ whose residues are a_2 and b_2 respectively, then we have relations similar to (18):

$$(21) \qquad t_1^{(2)} = t_1^{(1)}, \qquad t_2^{(2)} = \frac{t_2^{(1)} - \alpha_2 t_1^{(1)}}{t_1^{(1)}}, \qquad t_3^{(2)} = \frac{t_3^{(1)} - \beta_2 t_1^{(1)}}{t_1^{(1)}},$$

where $t_1^{(2)}$, $t_2^{(2)}$, $t_3^{(2)}$ are uniformizing parameters of $P^{(2)}(V^{(2)})$. More generally, we find:

$$(22)\quad t_1^{(i+1)} = t_1^{(i)}, \qquad t_2^{(i+1)} = \frac{t_2^{(i)} - \alpha_{i+1} t_1^{(i)}}{t_1^{(i)}}, \qquad t_3^{(i+1)} = \frac{t_3^{(i)} - \beta_{i+1} t_1^{(i)}}{t_1^{(i)}},$$

where $\alpha_{i+1}, \beta_{i+1} \in \mathfrak{J}$.

From (18), (21) and (22) we obtain the following relations:

$$(23)\quad \begin{aligned} t_2 &= \alpha_1 t_1 + \alpha_2 t_1^2 + \cdots + \alpha_i t_1^i + (\alpha_{i+1} + t_2^{(i+1)}) t_1^{i+1}, \\ t_3 &= \beta_1 t_1 + \beta_2 t_1^2 + \cdots + \beta_i t_1^i + (\beta_{i+1} + t_3^{(i+1)}) t_1^{i+1}. \end{aligned}$$

The relations (23) point obviously to the fact that the successive centers $P, P^{(1)}, P^{(2)}, \cdots$ of the valuation v arise from a sequence of "infinitely near" points on an algebraic or analytical branch γ on F, given by the expansions: $t_2 = \alpha_1 t_1 + \alpha_2 t_1^2 + \cdots$, $t_3 = \beta_1 t_1 + \beta_2 t_1^2 + \cdots$, and that the valuation v is therefore either algebraic (i.e. has rank 2) or analytical (i.e. has rank 1 and is discrete). The formal proof is as follows:

For simplicity of notation we agree to use one and the same symbol for an element of $\mathfrak{J}$ and for the F-residue of that element. Thus we shall treat t_1, t_2, t_3, if necessary, as elements of the field of rational functions on F, although in reality they are elements of $Q_V(P)$, i.e. elements of the field of rational functions on V. This convention permits us to speak of the values $v(t_1)$, $v(t_2)$, $v(t_2 - \alpha_1 t_1)$, etc. We shall adhere to this convention throughout this part of the paper.

Let

$$(24)\quad \begin{aligned} g_i &= t_2 - \alpha_1 t_1 - \alpha_2 t_1^2 - \cdots - \alpha_i t_1^i, \\ h_i &= t_3 - \beta_1 t_1 - \beta_2 t_1^2 - \cdots - \beta_i t_1^i, \end{aligned}$$

whence, by (23),

$$(25)\quad v(g_i) \geqq v(t_1^{i+1}), \qquad v(h_i) \geqq v(t_1^{i+1}).$$

Let ζ be an arbitrary element of $Q_F(P)$, and let $n_i(\zeta)$ be an integer defined as follows: a) if ζ, regarded as an element of $\mathfrak{J}$ belongs to the ideal $\mathfrak{J}\cdot(g_i, h_i)$, then $n_i(\zeta) = i + 1$; b) if ζ does not belong to the ideal $\mathfrak{J}\cdot(g_i, h_i)$, there exists an integer j (depending on ζ and i) such that

$$(26)\quad \zeta \equiv \epsilon \cdot t_1^j \ (\text{mod } \mathfrak{J}(g_i, h_i)),$$

where ϵ is a unit in $\mathfrak{J}$[6]. In this case we put

$$(27)\quad n_i(\zeta) = \min.\ (j, i + 1).$$

[6] The leading forms of g_i and h_i are respectively $z_2 - a_1 z_1$, $z_3 - b_1 z_1$. Hence t_1, g_i and h_i are uniformizing parameters of $P(V)$, and therefore the ideal $\mathfrak{J}\cdot(g_i, h_i)$ defines an algebraic curve Γ_i on V which has at P a simple point. If $\bar{\zeta}$ and $\bar{t}_1$ are the Γ_i-residues of ζ and t_1 respectively, then $\bar{\zeta} \neq 0$ and $\bar{t}_1$ is a uniformizing parameter of $P(\Gamma_i)$. Therefore we can write $\bar{\zeta} = \bar{\epsilon}\bar{t}_1^j$, where $\bar{\epsilon}$ is a unit in $Q_{\Gamma_i}(P)$. From this, congruence (26) follows immediately (ϵ = any element of $\mathfrak{J}$ whose Γ_i-residue is $\bar{\epsilon}$).

We have from (25) and (26): $v(\zeta) \geqq n_i(\zeta)v(t_1)$. If $n_i(\zeta) \to +\infty$ as $i \to +\infty$, then the valuation v is of rank 2 and is composed of a divisor whose center is a component of the curve defined on F by the principal ideal (ζ). Suppose now that v is not of rank 2. Then $n_i(\zeta)$ is bounded for any non zero element ζ in $Q_F(P)$. We conclude from (25), (26) and (27) that if i is sufficiently high we must have

$$v(\zeta) = n_i(\zeta)v(t_1),$$

whence *the value of any non zero element ζ in $Q_F(P)$ is an integer.* Therefore the valuation v is of rank 1. This completes the proof of the lemma.

12. Valuations of rank 2

The valuation v is composite with a divisor (a 1-dimensional valuation). This divisor may be of second kind with respect to F, i.e., its center may be the point P. It is known, however, that a finite number of quadratic transformations will necessarily lead to a surface $F^{(i)}$ with respect to which the divisor is of the first kind (see [5], p. 681; the reasoning is identical with that employed in the proof of Lemma 9.1). Hence it is permissible to assume that the center of the divisor is an irreducible curve Γ on F, through P. If Γ is a ν-fold curve of F (where ν is the multiplicity of the point P), then P is a simple point of Γ (see section 10). Hence successive monoidal transformations operating on Γ and on its transforms are permissible. By Theorem 4′ (section 9), a finite number of such transformations will eliminate Γ *as a ν-fold curve.* Hence we may assume that the multiplicity of Γ for F is less than ν (we assume of course that $\nu > 1$). *From now on we shall employ only quadratic transformations.*

First of all we may eliminate by quadratic transformations the singularity which Γ may possibly possess at P. Note that the successive centers $P^{(1)}, P^{(2)}, \cdots$ of the valuation v will always lie on successive transforms of the curve Γ, for the valuation v is composite with a divisor whose center is Γ. Hence it is permissible to assume that *P is a simple point of Γ.* We may then select uniformizing parameters x, y, z of $P(V)$ in such a fashion that the curve Γ be given by the equations $y = z = 0$ [i.e. Γ is given in $Q_V(P)$ by the ideal (y, z)]. Since $v(y) > nv(x)$ and $v(z) > nv(x)$ for any integer n, it follows that the local equations of our successive quadratic transformations $T^{(i)}$ will all be of the form:

$$x_{i+1} = x, \qquad y_{i+1} = \frac{y_i}{x}, \qquad z_{i+1} = \frac{z_i}{x},$$

where x, y_i, z_i are uniformizing parameters of $P^{(i)}(V^{(i)})$. On each surface $F^{(i)}$ the center of the divisor is the curve $y_i = z_i = 0$.

Let (ω_i) be the principal ideal in $\mathfrak{Z}_i$ [$= Q_{V^{(i)}}(P^{(i)})$] which defines the surface $F^{(i)}$. Here ω_{i+1} is the $T^{(i)}$-transform of ω_i (see section 11). As long as P_i is ν-fold for $F^{(i)}$ we will have

$$\omega_i = x^\nu \omega_{i+1}. \tag{28}$$

In particular, F is defined by the principal ideal $\mathfrak{Z}\cdot\omega$.

Let $\mathfrak{A}$ denote the ideal $\mathfrak{J}\cdot(y, z)$, and let $\mathfrak{m}_i$ be the ideal of non units in $\mathfrak{J}_i$.

LEMMA 12.1. *Let ξ be an element of $\mathfrak{J}$ which is exactly divisible by $\mathfrak{m}^\rho$ and let ξ_1 be the T-transform of ξ, i.e. let $\xi = x^\rho\xi_1$. A necessary condition that ξ_1 be divisible by $\mathfrak{m}_1^\sigma$ is that the following congruence be satisfied:*

$$\xi \equiv 0(\mathfrak{A}^\sigma \mathfrak{m}^{\rho-\sigma}, \mathfrak{m}^{\rho+1}). \tag{29}$$

PROOF. We shall use induction with respect to σ, since for $\sigma = 0$ the lemma is trivial. Assume that the lemma is true for $\sigma = s$ and let ξ_1 be divisible by $\mathfrak{m}_1^{s+1}$. Then ξ_1 is also divisible by $\mathfrak{m}_1^s$, (29) holds (with σ replaced by s), and we can write:

$$\xi = x^{\rho-s}\varphi_s(y, z) + \varphi_{s+1}(y, z) + \psi_{\rho+1}(x, y, z), \tag{30}$$

where φ_s is a form with coefficients in $\mathfrak{J}$, φ_{s+1} is a form with coefficients in $\mathfrak{m}^{\rho-s-1}$ and $\psi_{\rho+1}$ is a form with coefficients in $\mathfrak{J}$, the degrees of these forms being s, $s + 1$ and $\rho + 1$ respectively. We have:

$$\xi_1 = \varphi_s(y_1, z_1) + \overset{*}{\varphi}_{s+1}(y_1, z_1) + x\psi_{\rho+1}(1, y_1, z_1),$$

where $\overset{*}{\varphi}_{s+1}$ is a form, of degree $s + 1$, with coefficients in $\mathfrak{J}$. From this expression of ξ_1 it follows immediately that if $\xi_1 \equiv 0(\mathfrak{m}_1^{s+1})$ then the coefficients of φ_s must all belong to $\mathfrak{m}$. Hence $\varphi_s(y, z) \equiv 0(\mathfrak{m}^{s+1})$, and consequently, by (30), $\xi \equiv 0(\mathfrak{A}^{s+1}\mathfrak{m}^{\rho-s-1}, \mathfrak{m}^{\rho+1})$, q.e.d.

LEMMA 12.2. *If ρ, s and σ are non-negative integers, $s \geqq \sigma$, then*

$$(\mathfrak{A}^\rho, \mathfrak{m}^s):x^\sigma = (\mathfrak{A}^\rho, \mathfrak{m}^{s-\sigma}). \tag{31}$$

PROOF. It is sufficient to prove the lemma for $\sigma = 1$. Letting $\sigma = 1$, we shall use induction with respect to s, since if $s \leqq \rho$ then the lemma is trivial. Let ξ be any element of $(\mathfrak{A}^\rho, \mathfrak{m}^{s+1}):x$. Then by our induction, ξ is an element in $(\mathfrak{A}^\rho, \mathfrak{m}^{s-1})$, and we can write $\xi = \alpha_\rho + \pi$, when $\alpha_\rho \equiv 0(\mathfrak{A}^\rho)$ and $\pi \equiv 0(\mathfrak{m}^{s-1})$. We must have $\pi x \equiv 0(\mathfrak{A}^\rho, \mathfrak{m}^{s+1})$, i.e.

$$\pi x = \phi_\rho(y, z) + \pi', \tag{32}$$

where ϕ_ρ is a form, of degree ρ, with coefficients in $\mathfrak{J}$ and where $\pi' \equiv 0(\mathfrak{m}^{s+1})$. If $\pi \equiv 0(\mathfrak{m}^s)$, then $\xi \equiv 0(\mathfrak{A}^\rho, \mathfrak{m}^s)$, and there is nothing to prove. Suppose then that $\pi \not\equiv 0(\mathfrak{m}^s)$. In this case πx is exactly divisible by $\mathfrak{m}^s$, and the relation (32) is possible if and only if also $\phi_\rho(y, z)$ is in $\mathfrak{m}^s$ and if πx and $\phi_\rho(yz)$ have the same leading form, for $\pi' \equiv 0(\mathfrak{m}^{s+1})$. Now no term of the leading form of $\phi_\rho(y, z)$ is of degree less than ρ in y and z. Hence the same must be true of the leading form of π. This implies that $\pi \equiv 0(\mathfrak{A}^\rho, \mathfrak{m}^s)$, and consequently also $\xi \equiv 0(\mathfrak{A}^\rho, \mathfrak{m}^s)$, q.e.d.

By means of the preceding two lemmas the proof of the local reduction theorem for the valuation v is readily completed. If $P^{(1)}$ is a ν-fold point of $F^{(1)}$, the leading form of ω must be a binary form in y, z, with coefficients in the residue field of P. Hence $\omega \equiv 0(\mathfrak{A}^\nu, \mathfrak{m}^{\nu+1})$. On the other hand, the curve Γ $(y = z = 0)$ is

not ν-fold for the surface F, i.e. $\omega \not\equiv 0(\mathfrak{A}^\nu)$. There exists therefore an integer $n \geqq 1$ such that[7]

$$\omega \equiv 0(\mathfrak{A}^\nu, \mathfrak{m}^{n+\nu}), \qquad \omega \not\equiv 0(\mathfrak{A}^\nu, \mathfrak{m}^{n+\nu+1}),$$

and we can write: $\omega = \phi_\nu(y, z) + \xi$, where $\xi \equiv 0(\mathfrak{m}^{n+\nu})$ and ϕ_ν is a form of degree ν, with coefficients in $\mathfrak{J}$. By (28) we have: $\omega_1 = \phi_\nu(y_1, z_1) + x^n\xi_1$, $\xi_1 \in \mathfrak{J}_1$. *We assert that* $\omega_1 \not\equiv 0(\mathfrak{A}_1^\nu, \mathfrak{m}_1^{n+\nu})$, where $\mathfrak{A}_1 = \mathfrak{J}_1 \cdot (y_1, z_1)$. For in the contrary case we would have: $x^n\xi_1 \equiv 0(\mathfrak{A}_1^\nu, \mathfrak{m}_1^{n+\nu})$, and hence, by Lemma 12.2 (applied to $\mathfrak{A}_1$ and $\mathfrak{m}_1$), $\xi_1 \equiv 0(\mathfrak{A}_1^\nu, \mathfrak{m}_1^\nu)$, i.e. $\xi_1 \equiv 0(\mathfrak{m}_1^\nu)$. It then follows from Lemma 12.1 (in which we now have $\rho = n + \nu$, $\sigma = \nu$) that $\xi \equiv 0(\mathfrak{A}^\nu\mathfrak{m}^n, \mathfrak{m}^{n+\nu+1})$, and consequently $\omega \equiv 0(\mathfrak{A}^\nu, \mathfrak{m}^{n+\nu+1})$. This is in contradiction with our definition of the integer n.

We thus conclude that if $P^{(1)}$ is still ν-fold for $F^{(1)}$, and if n_1 is the integer with the property

$$\omega_1 \equiv 0(\mathfrak{A}_1^\nu, \mathfrak{m}_1^{n_1+\nu}), \qquad \omega_1 \not\equiv 0(\mathfrak{A}_1^\nu, \mathfrak{m}_1^{n_1+\nu+1}),$$

then

$$n > n_1 \geqq 0.$$

More generally, if $P^{(1)}, P^{(2)}, \cdots, P^{(i)}$ are ν-fold points of $F^{(1)}, F^{(2)}, \cdots, F^{(i)}$, and if n_i is the integer with the property

$$\omega_i \equiv 0(\mathfrak{A}_i^\nu, m_i^{n_i+\nu}), \qquad \omega_i \not\equiv 0(\mathfrak{A}_i^\nu, m_i^{n_i+\nu+1}),$$

where $\mathfrak{A}_i = \mathfrak{J}_i \cdot (y_i, z_i)$, then

$$n > n_1 > n_2 > \cdots > n_i \geqq 0.$$

This shows that the number of successive ν-fold points $P^{(i)}$ is necessarily finite. This completes the proof of the local reduction theorem for valuations of rank 2.

13. Differentiation with respect to uniformizing parameters

In this and in the remaining sections of Part II we assume that the ground field k is of characteristic zero. Let P be a simple point of an r-dimensional variety V and let $t_1, t_2, \cdots, t_r$ be uniformizing parameters of $P(V)$. If $f(z_1, z_2, \cdots, z_r)$ is a non-zero polynomial in the indeterminates z_i, with coefficients in k, and if $f_\rho(z_1, z_2, \cdots, z_r)$ is the leading form of f, i.e. the sum of terms of lowest degree ρ, then $f_\rho(z_1, z_2, \cdots, z_r)$ is also the leading form of $f(t_1, t_2, \cdots, t_r)$. Hence $f(t_1, t_2, \cdots, t_r)$ is different from zero since it has a leading form which is different from zero. We conclude that *the uniformizing parameters $t_1, t_2, \cdots, t_r$ are algebraically independent over k.* Hence every element ω of the field of rational functions on V is algebraic over the field $k(t_1, t_2, \cdots, t_r)$, and the partial derivatives $\partial\omega/\partial t_\alpha$, $\alpha = 1, 2, \cdots, r$, are defined and are elements of the field.

LEMMA 13.1. *If $\omega \in Q_V(P)$, then also $\dfrac{\partial\omega}{\partial t_\alpha} \in Q_V(P)$, $\alpha = 1, 2, \cdots, r$.*

---

[7] See, for instance, Krull [1], p. 207.

PROOF. Let $\xi_1, \xi_2, \cdots, \xi_n$ be a set of non-homogeneous coördinates of V such that P is at finite distance with respect to this system of coördinates. Since every element in $Q_V(P)$ is a quotient of polynomials in $k[\xi_1, \xi_2, \cdots, \xi_n]$ in which the denominator is $\neq 0$ at P, to prove the lemma it is sufficient to prove that the derivatives $\partial\xi_i/\partial t_\alpha$ belong to $Q_V(P)$.

Let $g_i(u)$ be an irreducible polynomial in $k[u]$ such that $g_i(\xi_i) = 0$ at P. Then $g_i(\xi_i)$ is a non-unit in $Q_V(P)$, and therefore, by the very definition of uniformizing parameters, we must have relations of the form:

$$H_i(\xi, t) = A(\xi)g_i(\xi_i) + \sum_{\alpha=1}^{r} B_{i\alpha}(\xi)t_\alpha = 0, \tag{33}$$

where ξ and t stand for the sets of elements $\xi_1, \xi_2, \cdots, \xi_n$ and $t_1, t_2, \cdots, t_r$ respectively. Here A and the $B_{i\alpha}$ are polynomials with coefficients in k and $A(\xi) \neq 0$ at P. Since $H_i(\xi, t) = 0$, we have

$$\frac{\partial H_i}{\partial t_\alpha} + \sum_{j=1}^{n} \frac{\partial H_i}{\partial \xi_j}\cdot\frac{\partial \xi_j}{\partial t_\alpha} = 0, \qquad \begin{array}{l} i = 1, 2, \cdots, n \\ \alpha = 1, 2, \cdots, r. \end{array} \tag{33'}$$

On the other hand, since $g_i'(\xi_i) \neq 0$ at P, we see from (33) that

$$\frac{\partial H_i}{\partial \xi_j} = 0 \quad \text{at } P, \text{ if } i \neq j,$$

$$\frac{\partial H_i}{\partial \xi_i} \neq 0 \quad \text{at } P.$$

Hence the determinant $\left|\frac{\partial H_i}{\partial \xi_j}\right|$ is different from zero at P, and consequently $\frac{\partial \xi_i}{\partial t_\alpha} \in Q_V(P)$, in view of (33′), q.e.d.

The following is an immediate consequence of the lemma. Let ω be an element of $Q_V(P)$ and let $F(z)$ be the leading form of ω, where F is of degree ν in $z_1, z_2, \cdots, z_r$. *Then* $\frac{\partial F}{\partial z_i}$ *is the leading form of* $\frac{\partial \omega}{\partial t_i}$, *unless* $\frac{\partial F}{\partial z_i} = 0$, *in which case* $\frac{\partial \omega}{\partial t_i} \equiv 0(\mathfrak{m}^\nu)$, *where* $\mathfrak{m}$ *is the ideal of non-units in* $Q_V(P)$. For if $F(z_1, z_2, z_3) = \sum_{(i)} a_{(i)} z_1^{i_1} z_2^{i_2} \cdots z_r^{i_r}$, then we can write ω in the form: $\omega = \sum_{(i)} \alpha_{(i)} t_1^{i_1} t_2^{i_2} \cdots t_r^{i_r}$, where $\alpha_{(i)}$ is some element of $Q_V(P)$ whose P-residue is $a_{(i)}$. Since $\partial\alpha_{(i)}/\partial t_j \in Q_V(P)$, it follows that $\frac{\partial \omega}{\partial t_1} \equiv \sum i_1 \alpha_{(i)} t_1^{i_1-1} t_2^{i_2} \cdots t_r^{i_r} \pmod{\mathfrak{m}^\nu}$, and from this our assertion follows. As a corollary we have the following familiar result: *if ω is exactly divisible by $\mathfrak{m}^\nu$, then all partial derivatives of ω, of order $1, 2, \cdots, \nu - 1$, with respect to the uniformizing parameters, are zero at P, but at least one partial derivative of order ν is different from zero at P; and conversely.*

In the following considerations we shall denote by z one of the uniformizing parameters of $P(V)$, say the parameter t_r. Thus in the case $r = 3$, which is of

particular interest to us, the uniformizing parameters will be denoted by x, y, z, as in the preceding section.

We introduce in $\mathfrak{J}$ [$= Q_V(P)$] an operator Δ_n, n — a positive integer, defined as follows: if ξ is any element of $Q_V(P)$, then

$$\Delta_n(\xi) = \xi - z\frac{\partial \xi}{\partial z} + \frac{z^2}{2!}\cdot\frac{\partial^2 \xi}{\partial z^2} - \cdots + (-1)^n \frac{z^n}{n!}\cdot\frac{\partial^n \xi}{\partial z^n}. \tag{34}$$

We have

$$\frac{\partial \Delta_n(\xi)}{\partial z} = (-1)^n \frac{z^n}{n!}\cdot\frac{\partial^{n+1}\xi}{\partial z^{n+1}} \equiv 0(\mathfrak{J}\cdot z^n). \tag{35}$$

Given any element ω in $\mathfrak{J}$ and given any positive integer ν, it is always possible to write ω in the form

$$\omega = \alpha_0 z^\nu + \alpha_1 z^{\nu-1} + \cdots + \alpha_{\nu-1} z + \alpha_\nu, \tag{36}$$

where $\alpha_i \in \mathfrak{J}$. Moreover, if $\omega \equiv 0(\mathfrak{m}^\nu)$, the α's can be so selected that they satisfy the congruences:

$$\alpha_i \equiv 0(\mathfrak{m}^i), \qquad i = 0, 1, \cdots, \nu. \tag{37}$$

Definition 1. The expression (36) of ω is *normal* if the elements α_i satisfy the congruences

$$\frac{\partial \alpha_i}{\partial z} \equiv 0(z^i), \qquad i = 0, 1, \cdots, \nu. \tag{38}$$

Definition 2. If $\omega \equiv 0(\mathfrak{m}^\nu)$, then the expression (36) of ω is *strongly normal* if the elements α_i satisfy the following condition: if α_i is exactly divisible by $\mathfrak{m}^{\nu_i}$, then $\nu_i \geqq i$ and

$$\frac{\partial \alpha_i}{\partial z} \equiv 0(z^{\nu_i}), \qquad i = 0, 1, \cdots, \nu. \tag{39}$$

We shall now prove the following lemma:

Lemma 13.2. *Any element ω of $\mathfrak{J}$ can be expressed in a normal form (for any integer ν). If $\omega \equiv 0(\mathfrak{m}^\nu)$ then the elements α_i which occur in a normal expression of ω satisfy the congruences* (37). *Any element ω in $\mathfrak{m}^\nu$ can be expressed in a strongly normal form.*

Proof. We start from an arbitrary expression of ω, of the form (36). By (34) we have $\Delta_n(\alpha_\nu) \equiv \alpha_\nu (\text{mod } z)$. Hence, if we put $\alpha'_\nu = \Delta_n(\alpha_\nu) = \alpha_\nu + \beta_{\nu-1}z$, then

$$\omega = \alpha_0 z^\nu + \alpha_1 z^{\nu-1} + \cdots + \alpha_{\nu-2}z^2 + (\alpha_{\nu-1} - \beta_{\nu-1})z + \alpha'_\nu.$$

This expression of ω is similar to (36), and we have $\dfrac{\partial \alpha'_\nu}{\partial z} \equiv 0(z^n)$, by (35). Hence if $n \geqq \nu$, then (38) is satisfied for $i = \nu$. Suppose that the congruences (38)

are satisfied for $i = s + 1, s + 2, \cdots \nu$. Then we put $\alpha'_s = \Delta_n(\alpha_s) = \alpha_s + \beta_{s-1}z$, and we replace the expression (36) of ω by the following:

$$\omega = \alpha_0 z^\nu + \alpha_1 z^{\nu-1} + \cdots + (\alpha_{s-1} - \beta_{s-1})z^{\nu-s+1} + \alpha'_s z^{\nu-s} + \alpha_{s+1}z^{\nu-s-1} + \cdots + \alpha_\nu,$$

and for this new expression of ω the congruences (38) are satisfied for $i = s$, $s + 1, \cdots, \nu$. Ultimately we get in this fashion a normal expression of ω.

Let us suppose now that (36) is a normal expression of ω, and let $\omega \equiv 0(\mathfrak{m}^\nu)$. All the partial derivatives of ω with respect to the uniformizing parameters $x, y, \cdots, z$, of order $\leqq \nu - 1$, vanish at P. Therefore all the partial derivatives of α_ν, of order $\leqq \nu - 1$, with respect to the parameters $x, y, \cdots$, other than z, must be zero at P. Since $\dfrac{\partial \alpha_\nu}{\partial z} \equiv 0(z^\nu)$, and since α_ν is certainly zero at P, it follows that $\alpha_\nu \equiv 0(\mathfrak{m}^\nu)$. Consequently

$$\omega' = \alpha_0 z^{\nu-1} + \alpha_1 z^{\nu-2} + \cdots + \alpha_{\nu-1} \equiv 0(\mathfrak{m}^{\nu-1}),$$

and since this is obviously a normal expression of ω', we conclude in a similar fashion that $\alpha_{\nu-1} \equiv 0(\mathfrak{m}^{\nu-1})$. Thus the congruences (37) follow by induction with respect to ν.

To derive a strongly normal expression of the element ω, we start with an arbitrary normal expression (36) of ω, and we assume that α_i is exactly divisible by $\mathfrak{m}^{\nu_i}$. From the preceding part of the proof we know that the inequalities $\nu_i \geqq i$ are automatically satisfied, since, by hypothesis, $\omega \equiv 0(\mathfrak{m}^\nu)$. Let us assume that the congruences (39) are already satisfied for $i = s + 1$, $s + 2, \cdots, \nu$, $s \leqq \nu$. We consider the element α_s. If the leading form of α_s is divisible by z, then α_s can be written in the form: $\alpha_s = \beta_{s-1}z + \bar{\alpha}_s$, where $\beta_{s-1} \in \mathfrak{J}$ and $\bar{\alpha}_s \equiv 0(\mathfrak{m}^{\nu_s+1})$, whence

$$\alpha_s \equiv 0(z, \bar{\alpha}_s). \tag{40}$$

This leads to another expression of ω which differs from (36) in that α_{s-1} is replaced by $\alpha_{s-1} + \beta_{s-1}$ and α_s is replaced by $\bar{\alpha}_s$. This expression may not be normal, but can be rendered normal according to the procedure of the first part of the proof. In this procedure the coefficients $\alpha_{s+1}, \alpha_{s+2}, \cdots \alpha_\nu$ are not affected, and $\bar{\alpha}_s$ is replaced by $\alpha'_s = \Delta_n(\bar{\alpha}_s)$, where $n \geqq s$. By the definition of the operator Δ_n it follows that

$$\alpha'_s \equiv \bar{\alpha}_s(\mathfrak{J}\cdot z) \quad \text{and} \quad \alpha'_s \equiv 0(\mathfrak{m}^{\nu_s+1}),$$

whence, by (40),

$$\alpha_s \equiv 0(z, \alpha'_s) \equiv 0(z, \mathfrak{m}^{\nu_s+1}).$$

If the leading form of α'_s is still divisible by z, the above procedure leads to a new normal form of ω:

$$\omega = \alpha''_0 z^\nu + \alpha''_1 z^{\nu-1} + \cdots + \alpha''_s z^{\nu-s} + \alpha_{s+1}z^{\nu-s+1} + \cdots + \alpha_\nu,$$

and we will have

$$\alpha_s' \equiv 0(z, \alpha_s''), \qquad \alpha_s'' \equiv 0(\mathfrak{m}^{\nu_s+2}),$$

whence

$$\alpha_s \equiv 0(z, \alpha_s'') \equiv 0(z, \mathfrak{m}^{\nu_s+2}).$$

If that process continues indefinitely, so that the leading form of $\alpha_s^{(i)}$ is always divisible by z, then $\alpha_s \equiv 0(z, \mathfrak{m}^{\nu_s+i})$, for all i, whence $\alpha_s \equiv 0(z)$, say $\alpha_s = z\beta_{s-1}$. In that case we get a normal form of ω:

$$\omega = \alpha_0 z^\nu + \alpha_1 z^{\nu-1} + \cdots + (\alpha_{s-1} + \beta_s)z^{\nu-s+1} + \alpha_{s+1}z^{\nu-s-1} + \cdots + \alpha_\nu$$

in which the coefficient of $z^{\nu-s}$ is zero. But if $\alpha_s = 0$, then the congruences (39) are satisfied for $i = s, s+1, \cdots, \nu$.

We may therefore assume that the leading form of α_s is not divisible by z. As was pointed out above, we may replace the original normal expression (36) of ω by another normal expression, in which α_s is replaced by $\Delta_n(\alpha_s)$, without affecting the coefficients $\alpha_{s+1}, \alpha_{s+2}, \cdots, \alpha_n$. Since $\alpha_s \equiv 0(\mathfrak{m}^{\nu_s})$, each of the $n + 1$ terms in the expression of $\Delta_n(\alpha_s)$, except the first (i.e. α_s), is in $\mathfrak{m}^{\nu_s}$, and its leading form *is* divisible by z, while α_s is *exactly* divisible by $\mathfrak{m}^{\nu_s}$ and its leading form is *not* divisible by z. Hence *$\Delta_n(\alpha_s)$ is still exactly divisible by $\mathfrak{m}^{\nu_s}$*. It is therefore sufficient to take $n \geqq \nu_s$ in order to satisfy the congruence $\partial\alpha_s/\partial z \equiv 0(z^{\nu_s})$ [see (35)]. This completes the proof of the lemma.

We conclude this section with two lemmas concerning the effect of a quadratic transformation on a normal form of an element ω. Let T be a quadratic transformation of V, of center P, and let P' be one of the points of the transform V' of V which correspond to V. *We assume that the quotients $t_2/t_1, t_3/t_1, \cdots, t_r/t_1$ are finite at P' and that, in particular, t_r/t_1, i.e. z/t_1, is zero at P'.* Then $z' = z/t_1$ is one of the uniformizing parameters of $P'(V')$. We consider a normal expression (36) of an element ω in $\mathfrak{J}$ and we assume that ω is exactly divisible by $\mathfrak{m}^\nu$. Then $\omega = t_1^\nu\omega'$, where ω' is the T-transform of ω. Moreover, by (37), we can write: $\alpha_i = t_1^i\alpha_i'$, $\alpha_i' \in \mathfrak{J}' = Q_{V'}(P')$, and therefore:

$$\omega' = \alpha_0'z'^\nu + \alpha_1'z'^{\nu-1} + \cdots + \alpha_{\nu-1}'z' + \alpha_\nu'. \tag{41}$$

LEMMA 13.3. *If* (36) *is a normal expression of ω, then* (41) *is likewise a normal expression of ω'.*

PROOF. We have, by hypothesis, $\dfrac{\partial\alpha_i}{\partial z} \equiv 0(z^i)$. We also have:

$$t_1^i\frac{\partial\alpha_i'}{\partial z'} = \frac{\partial\alpha_i}{\partial z'} = \frac{\partial\alpha_i}{\partial z}\cdot\frac{\partial z}{\partial z'} = t_1\frac{\partial\alpha_i}{\partial z} \equiv 0(z^i) \equiv 0(t_1^iz'^i).$$

Hence $\partial\alpha_i'/\partial z' \equiv 0(z'^i)$, q.e.d.

The principal ideal $\mathfrak{J}\cdot z$ defines an irreducible $(r-1)$-dimensional subvariety of V which has at P a simple point. The uniformizing parameters of $P(W)$ are $t_1, t_2, \cdots, t_{r-1}$ (more precisely: the residues of $t_1, t_2, \cdots, t_{r-1}$ mod $\mathfrak{J}\cdot t_r$). Under our assumption that the quotients t_i/t_1 are finite at P' and that z/t_1

is zero at P', it is clear that the T-transform of W is a variety W' passing through P', and that W' is defined in $\mathfrak{Z}'$ $[= Q_{V'}(P')]$ by the principal ideal $\mathfrak{Z}' \cdot z'$. The transformation $\overline{T}$ from W to W' which is induced by T is quadratic with center at P. Let $\mathfrak{m}$ denote the ideal of non units in $Q_W(P)$ and let $\mathfrak{m}'$ denote the ideals of non units in $Q_{V'}(P')$.

LEMMA 13.4. *If α is an element of $\mathfrak{Z}$ such that the congruence $\alpha \equiv 0(\mathfrak{m}^\rho)$ always implies the congruence $\partial\alpha/\partial z \equiv 0(\mathfrak{Z}\cdot z^\rho)$, then the T-transform α' of α satisfies the same condition (with respect to $\mathfrak{m}'$, $\mathfrak{Z}'$ and z'). Moreover, if α is exactly divisible by $\mathfrak{m}^r$, and if $\bar{\alpha}$ is the W-residue of α, then also $\bar{\alpha}$ is exactly divisible by $\overline{\mathfrak{m}}^r$, and the $\overline{T}$-transform of $\bar{\alpha}$ is the W'-residue of the T-transform of α.*

PROOF. If α is exactly divisible by $\mathfrak{m}^r$, then α' is at most divisible by $\mathfrak{m}'^r$, (Lemma 3.2), while from the proof of the preceding lemma it follows that the congruence $\partial\alpha/\partial z \equiv 0(z^r)$ implies the congruence $\partial\alpha'/\partial z' \equiv 0(z'^r)$.

Since α is exactly divisible by $\mathfrak{m}^r$ and since $\partial\alpha/\partial z \equiv 0(z^r) \equiv 0(\mathfrak{m}^r)$, it follows that the leading form of α must be independent of z. Hence the leading form of α is not divisible by z and consequently $\alpha \not\equiv 0(z, \mathfrak{m}^{r+1})$. This implies that $\bar{\alpha} \not\equiv 0(\overline{\mathfrak{m}}^{r+1})$. Hence $\bar{\alpha}$ is exactly divisible by $\overline{\mathfrak{m}}^r$.

We have therefore: $\alpha = t_1^r\alpha'$, $\bar{\alpha} = \bar{t}_1^r\bar{\alpha}'$, where $\bar{t}_1$ is the W'-residue of t_1 and where $\bar{\alpha}'$ is the $\overline{T}$-transform of $\bar{\alpha}$. From this it follows that $\bar{\alpha}'$ is the W'-residue of α', and this completes the proof of the lemma.

14. Valuations of rank 1

We now proceed to prove the local reduction theorem for any zero-dimensional valuation v of rank 1. We use the notations of section 10. If it is possible to lower the multiplicity ν of P by using quadratic transformations only, then there is nothing to prove. *We shall therefore assume that it is not possible to lower the multiplicity of P by using only quadratic transformations of the ambient variety V.* If we then denote by

$$F,\ F^{(1)},\ F^{(2)},\ \cdots,\ F^{(i)},\ \cdots$$

the successive quadratic transforms of F which are *determined* by the valuation v, then our assumption implies that the centers P, $P^{(1)}$, $P^{(2)}$, $\cdots$ of v on the surfaces F, $F^{(1)}$, $F^{(2)}$, $\cdots$ are all ν-fold points. Let

$$\mathfrak{Z}_i = Q_{V^{(i)}}(P^{(i)}), \qquad \mathfrak{m}^{(i)} = \text{ideal of non units in } \mathfrak{Z}_i,$$

and let $\mathfrak{Z}_i \cdot \omega_i$ denote the principal ideal which defines the surface $F^{(i)}$. Here ω_{i+1} is the $T^{(i)}$-transform of ω_i.

We begin by observing that *the leading form of ω_i is necessarily the ν^{th} power of a linear form, for all i.* For suppose that the contrary is true. Then we may assume that the leading form of ω itself ($i = 0$) is not a ν^{th} power. In that case the valuation v is discrete, by Lemma 11.4. Moreover, the leading form of no ω_i is a ν^{th} power [see (20) and the statement which immediately follows that formula], and the residue fields of the consecutive centers P, $P^{(1)}$, $P^{(2)}$, $\cdots$, all coincide (Theorem 3, section 7). Under these conditions, our assumption

that all points $P^{(i)}$ are ν-fold, $\nu > 1$, is impossible. This has been proved by us in [5], p. 651–652, under the hypothesis that the ground field is algebraically closed. But it is clear from the proof that what really matters is that the points $P^{(i)}$ all have the same residue field, and this condition is satisfied in the present case.

Let x, y, z be uniformizing parameters of $P(V)$. We assume that z actually occurs in the leading form of ω. If z' is any other element of $\mathfrak{J}$ such that also x, y, z' are uniformizing parameters of $P(V)$, then it is clear that also z' will actually occur in the leading form of ω. We shall select the element z so as to satisfy a certain condition.

LEMMA 14.1. *The parameters x and y being fixed, there exists a third uniformizing parameter z such that the following condition is satisfied: if z' is any element of $\mathfrak{J}$ such that x, y, z' are also uniformizing parameters of $P(V)$, then the relations:*

$$v(z') > v(z), \tag{42}$$

$$v\left(\frac{\partial \omega}{\partial z}\right) \geqq v(z) \tag{42'}$$

are never satisfied simultaneously.

PROOF. Suppose that the parameter z does not satisfy the required condition. Then (42′) holds and there exists another parameter z' for which (42) holds. If we write z' in the form $z' = \alpha x + \beta y + \gamma z$, where α, β, γ are elements of $\mathfrak{J}$, then the hypothesis that also x, y and z' are uniformizing parameters of $P(V)$ implies that γ is a unit in $\mathfrak{J}$, whence $v(\gamma) = 0$. Since

$$\partial z'/\partial z = \gamma + \frac{\partial \alpha}{\partial z} x + \frac{\partial \beta}{\partial z} y + \frac{\partial \gamma}{\partial z} z,$$

it follows that also $\partial z'/\partial z$ is a unit in $\mathfrak{J}$. Consequently

$$v\left(\frac{\partial \omega}{\partial z}\right) = v\left(\frac{\partial \omega}{\partial z'} \cdot \frac{\partial z'}{\partial z}\right) = v\left(\frac{\partial \omega}{\partial z'}\right).$$

If also z' does not satisfy the required condition, then

$$v\left(\frac{\partial \omega}{\partial z}\right) = v\left(\frac{\partial \omega}{\partial z'}\right) \geqq v(z'),$$

and there will exist another parameter z'' such that

$$v(z'') > v(z').$$

If this continues indefinitely, then we get a sequence of elements $z, z', z'', \cdots, z^{(i)}, \cdots$ such that

$$v(z) < v(z') < v(z'') < \cdots < v(z^i) < \cdots \tag{43}$$

$$v\left(\frac{\partial \omega}{\partial z}\right) = v\left(\frac{\partial \omega}{\partial z^i}\right) \geqq v(z^{(i)}). \tag{43'}$$

From the fact that the valuation v is of rank 1 and that every ideal in $\mathfrak{J}$ has a finite base, follows immediately that the values assumed by the elements of $\mathfrak{J}$, if arranged in order of magnitude, form a simple sequence whose limit is $+\infty$[8]. Consequently, by (43), we have $\lim. v(z^{(i)}) = +\infty$, whence, by (43′), $\partial\omega/\partial z = 0$ on F. Hence $\partial\omega/\partial z$ must be divisible by ω (in $\mathfrak{J}$). This is impossible, since the leading form of ω actually involves z and therefore $\partial\omega/\partial z$ is exactly divisible by $\mathfrak{m}^{\nu-1}$, while ω is divisible by $\mathfrak{m}^{\nu}$. This contradiction completes the proof of the lemma.

From now on we assume that z satisfies the condition of Lemma 14.1.

We denote by $v(\mathfrak{m})$ the least value assumed by elements of $\mathfrak{m}$. It is clear that $v(\mathfrak{m}) = \text{minimum}\,[v(x), v(y), v(z)]$. Since the leading form of ω actually involves z, either the ratios y/x, z/x, or the ratios x/y, z/y are finite at $P^{(1)}$; in other words: either $v(x) = v(\mathfrak{m})$ or $v(y) = v(\mathfrak{m})$. We fix our notation so that y/x and z/x are finite at $P^{(1)}$, whence

$$v(x) = v(\mathfrak{m}). \tag{44}$$

We shall prove in a moment that $v(z) > v(\mathfrak{m})$ (Lemma 14.2). Assuming this, we see that z/x is zero at $P^{(1)}$, whence x and $z_1 = z/x$ can be taken as two of a set of three uniformizing parameters of $P^{(1)}(V^{(1)})$, while as third uniformizing parameter we can take an element of the form $f(y/x)$, where f is a polynomial with coefficients in $\mathfrak{J}$. The polynomial $f(t)$ must satisfy the condition that if its coefficients are reduced modulo $\mathfrak{m}$, the resulting polynomial $F(t)$ (with coefficients in the residue field of P) is irreducible and vanishes for $t = t_0 = P^{(1)}$-residue of y/x (all this follows from section 2). Thus the uniformizing parameters of $P^{(1)}(V^{(1)})$ are x, $f(y/x)$ and z_1. Let the pair of parameters $(x, f(y/x))$ be denoted by (x_1, y_1), where we fix our notation so that $v(x_1) \leqq v(y_1)$ (so that x_1 is not necessarily the element x). Since $v(z_1) \geqq v(\mathfrak{m}^{(1)})$, and since $z = xz_1$, it follows from (44) that $v(z) \geqq v(\mathfrak{m}) + v(\mathfrak{m}^{(1)})$. We have thus shown that if $v(z) > v(\mathfrak{m})$, then $v(z) \geqq v(\mathfrak{m}) + v(\mathfrak{m}^{(1)})$.

Since the leading form of ω involves z, the leading form of ω_1 will actually involve z_1 [see formula (19), section 11]. Hence we may assert again that y_1/x_1 and z_1/x_1 are finite at $P^{(2)}$, i.e. that $v(x_1) = v(\mathfrak{m}^{(1)})$. *The Lemma* 14.2 *will show that again we must have* $v(z_1) > v(x_1)$, i.e. $v(z_1) > v(\mathfrak{m}^{(1)})$, and consequently we find, as before, uniformizing parameters x_2, y_2, z_2 of $P^{(2)}(V^{(2)})$, where $z_2 = z_1/x_1$. Here $v(x_2) \leqq v(y_2)$ and one of the two elements x_2, y_2 coincides with x_1. The other element is of the form $f_1(y_1/x_1)$, where f_1 is a polynomial with coefficients in $\mathfrak{J}_1$. The relation $v(z_1) > v(x_1)$ is equivalent to the relation $v(z) > v(m) + v(\mathfrak{m}^{(1)})$, and again we point out that this relation implies the following: $v(z) \geqq v(\mathfrak{m}) + v(\mathfrak{m}^{(1)}) + v(\mathfrak{m}^{(2)})$ (since $v(z_2) \geqq v(\mathfrak{m}^{(2)})$).

[8] Let α be an arbitrary element of $\mathfrak{J}$ and let $\mathfrak{A}$ be the ideal in $\mathfrak{J}$ whose elements have value $\geqq v(\alpha)$. Since all elements of $\mathfrak{m}$ have positive value and since v is of rank 1, we have $\mathfrak{m}^{\rho} \equiv 0(\mathfrak{A})$ for ρ sufficiently high. Since $\mathfrak{m}^{\rho}$ has finite length it follows that if the values $v(\alpha)$, $\alpha \in \mathfrak{J}$, are arranged in order of magnitude, then each $v(\alpha)$ is preceded by a finite number of values $v(\beta)$, $\beta \in \mathfrak{J}$. Since $\lim_{\rho\to+\infty} v(\mathfrak{m}^{\rho}) = +\infty$, our assertion follows.

The possibility of continuing the above procedure indefinitely depends at each stage on the validity of the inequality: $v(z_i) > v(x_i) = v(\mathfrak{m}^{(i)})$, or of the equivalent inequality $v(z) > v(\mathfrak{m}) + v(\mathfrak{m}^{(1)}) + \cdots + v(\mathfrak{m}^{(i)})$. Hence we shall prove the following lemma:

LEMMA 14.2. *The inequality*

$$v(z) > v(\mathfrak{m}) + v(\mathfrak{m}^{(1)}) + \cdots + v(\mathfrak{m}^{(i)}) \tag{45}$$

holds for all values of i.

PROOF. Let us assume that the lemma is false and let us denote by $n - 2(n \geqq 1)$ the greatest value of i for which (45) holds true. At each of the points $P, P^{(1)}, P^{(2)}, \cdots, P^{(n-1)}$ we will have then uniformizing parameters (x_i, y_i, z_i), such that: 1) $z_i = z_{i-1}/x_{i-1}$; 2) $v(x_i) = v(\mathfrak{m}^{(i)})$; 3) one of the two elements x_i, y_i coincides with x_{i-1}, while the second element is of the form $f_{i-1}(y_{i-1}/x_{i-1})$, where f_{i-1} is a polynomial with coefficients in $\mathfrak{J}_{i-1}$. But while $v(z_i/x_i) > 0$, for $i = 0, 1, \cdots, n - 2$ $[(x_0, y_0, z_0) = (x, y, z)]$, we must have, by hypothesis, $v(z_{n-1}/x_{n-1}) = 0$, *whence* z_{n-1}/x_{n-1} *is a unit in* $Q_{V^{(n)}}(P^{(n)})$.

Let

$$\omega = \alpha_0 z^\nu + \alpha_1 z^{\nu-1} + \cdots + \alpha_\nu \tag{46}$$

be a normal expression for ω (see Definition 1, section 13). Then

$$\omega_1 = \alpha_0^{(1)} z_1^\nu + \alpha_1^{(1)} z_1^{\nu-1} + \cdots + \alpha_\nu^{(1)}, \tag{46.1}$$

where

$$\omega = x^\nu \omega_1, \qquad \alpha_j = x^j \alpha_j^{(1)}, \qquad \alpha_j^{(1)} \in \mathfrak{J}_1, \tag{47.1}$$

and the expression (46.1) for ω_1 is normal (Lemma 13.3). Continuing in this fashion we get a normal expression for each of the elements $\omega_1, \omega_2, \cdots, \omega_{n-1}$:

$$\omega_i = \alpha_0^{(i)} z_i^\nu + \alpha_1^{(i)} z_i^{\nu-1} + \cdots + \alpha_\nu^{(i)}, \qquad i = 1, 2, \cdots, n-1, \tag{46.i}$$

where

$$\omega_{i-1} = x_{i-1}^\nu \omega_i, \qquad \alpha_j^{(i-1)} = x_{i-1}^j \alpha_j^{(i)}, \qquad \alpha_j^{(i)} \in \mathfrak{J}_i. \tag{47.i}$$

We also have a similar expression for ω_n:

$$\omega_n = \alpha_0^{(n)} z_n^\nu + \alpha_1^{(n)} z_n^{\nu-1} + \cdots + \alpha_\nu^{(n)}, \tag{46.n}$$

where $z_n = z_{n-1}/x_{n-1}$ and

$$\omega_{n-1} = x_{n-1}^\nu \omega_n, \qquad \alpha_j^{(n-1)} = x_{n-1}^j \alpha_j^{(n)}, \qquad \alpha_j^{(n)} \in \mathfrak{J}_n, \tag{47.n}$$

but this time z_n *is a unit in* $\mathfrak{J}_n$.

Let c_j and d be the P_n-residues of $\alpha_j^{(n)}$ and of z_n respectively $(d \neq 0)$. Since $\omega_n \equiv 0(\mathfrak{m}^{(n)\nu})$, we must have, by (46.n):

$$c_0 u^\nu + c_1 u^{\nu-1} + \cdots + c_\nu = c_0(u - d)^\nu.$$

From this we deduce the following consequences:

a) All c_j are different from zero. Consequently the $\nu + 1$ terms $\alpha_j^{(n)} z_n^{\nu-j}$ have value zero in the valuation v. But since each of these terms differs from the corresponding term $\alpha_j z^{\nu-j}$ in (46) by the factor $x^\nu x_1^\nu \cdots x_{n-1}^\nu$, it follows that

$$v(\alpha_0 z^\nu) = v(\alpha_1 z^{\nu-1}) = \cdots = v(\alpha_\nu). \tag{48}$$

b) The residue d of z_n coincides with the residue $-c_1/\nu c_0$ of $-\alpha_1^{(n)}/\nu\alpha_0^{(n)}$, or—what is the same—the residue of $\alpha_1^{(n)}/\nu\alpha_0^{(n)} z_n$ equals -1. Since

$$\alpha_1^{(n)}/\alpha_0^{(n)} z_n = \alpha_1^{(n)} z_n^{\nu-1}/\alpha_0^{(n)} z_n^\nu = \alpha_1 z^{\nu-1}/\alpha_0 z^\nu = \alpha_1/\alpha_0 z,$$

we conclude that the residue of $\alpha_1/\nu\alpha_0 z$ in our valuation v is equal to -1. Taking into account that α_0 is a unit ($\alpha_0 = \alpha_0^{(n)}$ and $c_0 \neq 0$), whence $v(\alpha_0) = 0$, we conclude that

$$v\left(z + \frac{\alpha_1}{\nu\alpha_0}\right) > v(z).$$

Since (46) is a normal expression of ω, we have $\alpha_1 \equiv 0(\mathfrak{m})$ and $\dfrac{\partial\alpha_1}{\partial z} \equiv 0(\mathfrak{m})$. Hence either $\alpha_1 \equiv 0(\mathfrak{m}^2)$ or the leading form of α_1 is independent of z. Consequently, if we put

$$z' = z + \frac{\alpha_1}{\nu\alpha_0},$$

then x, y, z' are uniformizing parameters of $P(V)$, and

$$v(z') > v(z). \tag{49}$$

We consider $\partial\omega/\partial z$:

$$\frac{\partial\omega}{\partial z} = [\nu\alpha_0 z^{\nu-1} + (\nu - 1)\alpha_1 z^{\nu-2} + \cdots + \alpha_{\nu-1}] + \left[\frac{\partial\alpha_0}{\partial z} z^\nu + \frac{\partial\alpha_1}{\partial z} z^{\nu-1} + \cdots + \frac{\partial\alpha_\nu}{\partial z}\right].$$

All terms in the first parenthesis have the same value, equal to $v(z^{\nu-1})$. All terms in the second parenthesis are divisible by z^ν, since $\dfrac{\partial\alpha_i}{\partial z} \equiv 0(z^i)$. Hence $v(\partial\omega/\partial z) \geqq v(z^{\nu-1}) \geqq v(z)$, since $\nu > 1$. This equality and the inequality (49) are in contradiction with our hypothesis that z satisfies the condition of Lemma 14.1. This completes the proof of Lemma 14.2.

In view of Lemma 14.2, the parameters x_i, y_i, z_i as described above for $i = 0, 1, \cdots, n - 1$, are actually defined for all values of i, and so is the normal expression (46.i) of ω_i. For each point $P^{(i)}$ we consider the surface $W^{(i)}$ through $P^{(i)}$ defined on $V^{(i)}$ by the principal ideal $\mathfrak{Z}_i \cdot z_i$. As was pointed out in the preceding section in connection with Lemma 13.4, $W^{(i+1)}$ is the transform of $W^{(i)}$ by our quadratic transformation $T^{(i)}$, and x_i, y_i are uniformizing parameters

of $P^{(i)}(W^{(i)})$. If we now assume that the expression (46) of ω is not only normal but also strongly normal (Lemma 13.2) and if we take into account Lemmas 11.2, 11.3 and 13.4 (these lemmas have to be applied to $\alpha_0, \alpha_1, \cdots, \alpha_\nu$), we conclude that for a sufficiently high value s of i, the expression (46.i) of ω_i will take the following form:

$$(50)\quad \omega_s = \epsilon_0^{(s)} z_s^\nu + \epsilon_1^{(s)} x_s^{m_1} y_s^{n_1} z_s^{\nu-1} + \cdots + \epsilon_{\nu-1}^{(s)} x_s^{m_{\nu-1}} y_s^{n_{\nu-1}} z_s + \epsilon_\nu^{(s)} x_s^{m_\nu} y_s^{n_\nu},$$

where m_j, n_j are non-negative integers and where each $\epsilon_j^{(s)}$ is either a unit in $\mathfrak{J}_s$ or is zero (but $\epsilon_0^{(s)} = \alpha_0^{(s)} = \alpha_0 \neq 0$).

At this stage of the proof the following two cases must be considered separately:

FIRST CASE. *$v(x_s)$ and $v(y_s)$ are rationally dependent.*

SECOND CASE. *$v(x_s)$ and $v(y_s)$ are rationally independent.* In the second case we are definitely dealing with a valuation of rational rank 2. This case will be considered in the next section. Here we shall deal with the first case.

If $v(x_s)$ and $v(y_s)$ are rationally dependent, there will exist a least integer σ ($\sigma \geqq s$) such that $v(x_\sigma) = v(y_\sigma)$, while for any integer i, $s < i \leqq \sigma$, we will have either $x_i = x_{i-1}$, $y_i = y_{i-1}/x_{i-1}$, or $x_i = y_{i-1}/x_{i-1}$ and $y_i = x_{i-1}$. It follows therefore from (50), that ω_σ will have the following form:

$$(51)\quad \omega_\sigma = \epsilon_0 z_\sigma^\nu + \epsilon_1 x_\sigma^{\lambda_1} z_\sigma^{\nu-1} + \cdots + \epsilon_{\nu-1} x_\sigma^{\lambda_{\nu-1}} z_\sigma + \epsilon_\nu x_\sigma^{\lambda_\nu},$$

where each ϵ_j is either a unit in $\mathfrak{J}_\sigma$ or is zero.

Since $P^{(\sigma)}$ is a ν-fold point of $F^{(\sigma)}$, and since the ϵ's which are not zero are units in $\mathfrak{J}_\sigma$, we must have

$$\lambda_j - j \geqq 0$$

for all j such that $\epsilon_j \neq 0$. Hence the curve Δ defined by the ideal $\mathfrak{J}_\sigma(x_\sigma, z_\sigma)$ is a ν-fold curve of $F^{(\sigma)}$. We shall now apply to $F^{(\sigma)}$ a monoidal transformation M of center Δ. This transformation is permissible since $P^{(\sigma)}$ is a simple point of Δ [besides, we know in advance from the discussion in section 10 that the quadratic transformations which led from F to $F^{(\sigma)}$ could not possibly lead to ν-fold points other than those of types a), b) and c) described in section 10]. Let F_1 and V_1 be respectively the M-transforms of $F^{(\sigma)}$ and of $V^{(\sigma)}$, and let P_1 be the center of the valuation v on F_1.

We put $z_{\sigma 1} = z_\sigma / x_\sigma$ and

$$(51.1)\quad \omega_{\sigma 1} = \epsilon_0 z_{\sigma 1}^\nu + \epsilon_1 x_\sigma^{\lambda_1 - 1} z_{\sigma 1}^{\nu-1} + \cdots + \epsilon_{\nu-1} x_\sigma^{\lambda_{\nu-1} - \nu + 1} z_{\sigma 1} + \epsilon_\nu x_\sigma^{\lambda_\nu - \nu}.$$

Since ϵ_0 is a unit, the leading form of ω_σ involves z_σ, whence $z_{\sigma 1}$ is finite at P_1. The surface F_1 is defined in $Q_{V_1}(P_1)$ by the principal ideal $(\omega_{\sigma 1})$. *If $z_{\sigma 1}$ is zero at P_1*, then $x_\sigma, y_\sigma, z_{\sigma 1}$ are uniformizing parameters of $P_1(V_1)$ [see section 7, equations (10)]. *Note that $z_{\sigma 1}$ is certainly zero at P_1 if* $\lambda_j - j > 0$, for all j such that $\epsilon_j \neq 0$.

Under the hypothesis that $z_{\sigma 1}$ is zero at P_1, the expression (51.1) of $\omega_{\sigma 1}$ is similar to that of ω_σ (51). If P_1 is still ν-fold for F_1, then we must have $\lambda_j - 2j \geqq 0$, and the curve Δ_1 defined by the ideal $(x_\sigma, z_{\sigma 1})$ is ν-fold for F_1. We then

apply a second monoidal transformation M_1 of center Δ_1, getting new transforms F_2 and V_2. We then put

$$z_{\sigma 2} = z_{\sigma 1}/x_\sigma = z_\sigma/x_\sigma^2,$$

$$\omega_{\sigma 2} = \epsilon_0 z_{\sigma 2}^{\nu} + \epsilon_1 x_\sigma^{\lambda_1 - 2} z_{\sigma 2}^{\nu - 1} + \cdots + \epsilon_{\nu-1} x_\sigma^{\lambda_{\nu-1} - 2(\nu-1)} z_{\sigma 2} + \epsilon_\nu x_\sigma^{\lambda_\nu - 2\nu},$$

where again $z_{\sigma 2}$ is certainly finite at the center P_2 of v on F_2 and where F_2 is defined in $Q_{V_2}(P_2)$ by the principal ideal $(\omega_{\sigma 2})$. If $\lambda_j - 3j \geqq 0$, then the curve Δ_2 defined by the ideal $(x_\sigma, z_{\sigma 2})$ is ν-fold, and we can again apply a monoidal transformation.

Let h be the greatest integer such that $\lambda_j - hj \geqq 0$ for all j such that $\epsilon_j \neq 0$. Then the preceding considerations show that after h monoidal transformations with centers at successive ν-fold curves we will get a surface F_h on a V_h, on which the center of v is a point P_h, and which is defined in $Q_{V_h}(P_h)$ by the principal ideal $(\omega_{\sigma h})$, where

$$\omega_{\sigma h} = \epsilon_0 z_{\sigma h}^{\nu} + \epsilon_1 x_\sigma^{\lambda_1'} z_{\sigma h}^{\nu-1} + \cdots + \epsilon_{\nu-1} x_\sigma^{\lambda_{\nu-1}'} z_{\sigma h} + \epsilon_\nu x_\sigma^{\lambda_\nu'}. \tag{52}$$

Here

$$\lambda_j' = \lambda_j - hj,$$

and

$$z_{\sigma h} = z_\sigma / x_\sigma^h.$$

If $z_{\sigma h}$ is zero at P_h, then x_σ, y_σ, $z_{\sigma h}$ are uniformizing parameters of $P_h(V_h)$. By the definition of the integer h, we cannot have $\lambda_j' - j \geqq 0$ for all j such that $\epsilon_j \neq 0$. Hence in this case P_h is of multiplicity less than ν for F_h, and the proof is complete.

Suppose, however, that $z_{\sigma h} \neq 0$ at P_h, and let d be the residue of $z_{\sigma h}$ at P_h, $d \neq 0$. Then if $c_0, c_1, \cdots c_\nu$ are the P_h-residues of ϵ_0, $\epsilon_1 x_\sigma^{\lambda_1'}$, $\cdots$, $\epsilon_\nu x_\sigma^{\lambda_\nu'}$, *and if suppose that* P_h *is* ν*-fold for* F_h, then $u = d$ must be a ν-fold root of the polynomial

$$c_0 u^\nu + c_1 u^{\nu-1} + \cdots + c_\nu.$$

This implies that $c_j \neq 0$, for $j = 0, 1, 2, \cdots, \nu$, whence $\lambda_j' = 0$ and no ϵ_j is zero. Thus (52) becomes:

$$\omega_{\sigma h} = \epsilon_0 z_{\sigma h}^{\nu} + \epsilon_1 z_{\sigma h}^{\nu-1} + \cdots + \epsilon_{\nu-1} z_{\sigma h} + \epsilon_\nu,$$

and each of the $\nu + 1$ terms $\epsilon_j z_{\sigma h}^{\nu-j}$ is a unit, and therefore has value zero in the valuation v. Since these $v + 1$ terms are proportional to the $\nu + 1$ terms $\alpha_j z^{\nu-j}$ in (46), we find again the equations (48). Moreover, since $d = -c_1/\nu c_0 = P_h$-residue of $-\epsilon_1/\nu\epsilon_0$, it follows that the residue of z (in the valuation v) equals the residue of $-\alpha_1/\nu\alpha_0$. As in the proof of Lemma 14.2, so also here these conclusions lead to a contradiction with our original hypothesis that z satisfies the condition of Lemma 14.1. Therefore P_h cannot be a ν-fold point of F_h. This completes the proof of the local reduction theorem under the hypothesis that $v(x_s)$ and $v(y_s)$ are rationally dependent.

15. Valuations of rational rank 2

We now consider the case where $v(x_s)$ and $v(y_s)$ are rationally independent. Then also $v(x_i)$ and $v(y_i)$ will be rationally independent for all $i > s$ and we will have either

$$x_{i+1} = x_i, \qquad y_{i+1} = y_i/x_i, \qquad z_{i+1} = z_i/x_i,$$

if $v(x_i) < v(y_i/x_i)$, or

$$x_{i+1} = y_i/x_i, \qquad y_{i+1} = x_i, \qquad z_{i+1} = z_i/x_i,$$

if $v(y_i/x_i) < v(x_i)$, for all $i \geqq s$. Let $s < j_1 < j_2 \cdots$ be the sequence of integers defined as follows: $v(x_i) > v(y_i/x_i)$, if $i = j_\alpha - 1$, $i \geqq s$, and $v(x_i) < v(y_i/x_i)$ for all other values of i, $i \geqq s$. We combine the successive quadratic transformations $T^{(s)}, T^{(s+1)}, \cdots, T^{(j_\alpha - 1)}$ into one transformation, which we shall denote by C_α, and we put

$$X = x_s, \qquad Y = y_s, \qquad X_\alpha = x_{j_\alpha}, \qquad Y_\alpha = y_{j_\alpha}, \qquad \bar{z}_\alpha = z_{j_\alpha}.$$

If we put $v(Y)/v(X) = \tau$, then it is immediately seen ([5], p. 653) that

$$X = X_\alpha^{g_{\alpha-1}} Y_\alpha^{g_\alpha}, \qquad Y = X_\alpha^{f_{\alpha-1}} Y_\alpha^{f_\alpha}, \tag{53}$$

where f_α/g_α are the convergent fractions of the irrational number τ ($f_0 = 1$, $g_0 = 0$).

For the element ω_{j_α} we will have an expression similar to (50):

$$\bar{\omega}_\alpha = \bar{\epsilon}_0^{(\alpha)} \bar{z}_\alpha^{\nu} + \bar{\epsilon}_1^{(\alpha)} X_\alpha^{\bar{m}_1} Y_\alpha^{\bar{n}_1} \bar{z}_\alpha^{\nu-1} + \cdots + \bar{\epsilon}_{\nu-1}^{(\alpha)} X_\alpha^{\bar{m}_{\nu-1}} Y_\alpha^{\bar{n}_{\nu-1}} \bar{z}_\alpha + \bar{\epsilon}_\nu^{(\alpha)} X_\alpha^{\bar{m}_\nu} Y_\alpha^{\bar{n}_\nu}, \tag{54}$$

where we have put $\bar{\omega}_\alpha = \omega_{j_\alpha}$, $\epsilon^{(j_\alpha)} = \bar{\epsilon}^{(\alpha)}$, and where the integers $\bar{m}_j$, $\bar{n}_j$ depend on j and α. We denote the surface $F^{(j_\alpha)}$ and the point $P^{(j_\alpha)}$ respectively by $\bar{F}_\alpha$ and $\bar{P}_\alpha$.

If $\bar{m}_j \geqq j$, for $j = 1, 2, \cdots, \nu$, then the curve $X_\alpha = \bar{z}_\alpha = 0$ is a ν-fold curve of $\bar{F}_\alpha$, and we can apply to $\bar{F}_\alpha$ a monoidal transformation with center at that curve. Similarly, if $\bar{n}_j \geqq j$, $j = 1, 2, \cdots, \nu$, we apply a monoidal transformation with center at the curve $Y_\alpha = \bar{z}_\alpha = 0$. If both curves are ν-fold for $\bar{F}_\alpha$, then $\bar{P}_\alpha$ is a normal crossing of these curves, and we can apply both monoidal transformations, the order in which these transformations are applied being immaterial (Lemma 8.3).

More generally, if s is the greatest integer such that $\bar{m}_j - sj \geqq 0$, for all j, and if t is the greatest integer such that $\bar{n}_j - tj \geqq 0$, for all j, then we can apply to $\bar{F}_\alpha$ s consecutive monoidal transformations, with center at the ν-fold curve $X_\alpha = \bar{z}_\alpha = 0$ and at the $s - 1$ successive transforms of this curve (all these $s - 1$ transforms will be ν-fold curves), and t consecutive monoidal transformations with center at the ν-fold curve $Y_\alpha = \bar{z}_\alpha = 0$ and at the successive transforms of this curve. Let the new surface thus obtained be denoted by F_α, and let P_α be the center of v on F_α. If we put

$$\bar{z}_\alpha = X_\alpha^s Y_\alpha^t Z_\alpha,$$

then the surface F_α will be defined in $Q(P_\alpha)$ by the principal ideal (Ω_α), where

(55) $$\Omega_\alpha = \bar{\epsilon}_0^{(\alpha)} Z_\alpha^\nu + \bar{\epsilon}_1^{(\alpha)} X_\alpha^{\bar{M}_1} Y_\alpha^{\bar{N}_1} Z_\alpha^{\nu-1} + \cdots + \bar{\epsilon}_{\nu-1}^{(\alpha)} X_\alpha^{\bar{M}_{\nu-1}} Y_\alpha^{\bar{N}_{\nu-1}} Z_\alpha + \bar{\epsilon}_\nu^{(\alpha)} X_\alpha^{\bar{M}_\nu} Y_\alpha^{\bar{N}_\nu},$$

where the integers $\bar{M}_j$, $\bar{N}_j$ (which depend on j and α) will satisfy the inequalities:

(56) $$\text{minimum } (\bar{M}_j - j) < 0, \qquad \text{minimum } (\bar{N}_j - j) < 0.$$

The case where $Z_\alpha \neq 0$ at P_α is settled by means of considerations similar to those developed in the preceding section in connection with the element $z_{\sigma h}$ occurring in (52). It follows namely in this case that the hypothesis that P_α is a ν-fold point of F_α is in contradiction with the hypothesis that our original element z satisfies the condition of Lemma 14.1. Hence we may assume that $Z_\alpha = 0$ at P_α, for all α, whence X_α, Y_α and Z_α are uniformizing parameters of $P_\alpha(V_\alpha)$.

We have now transformed the surface $F^{(s)}$ into the surface F_α by a sequence of permissible transformations. If we put $z_s = Z$, then the uniformizing parameters of $P^{(s)}(V^{(s)})$ are X, Y, Z, and $F^{(s)}$ is given in $Q_{V^{(s)}}(P^{(s)})$ by the principal ideal (ω_s), where [see (50)],

(57) $$\omega_s = \epsilon_0^{(s)} Z^\nu + \epsilon_1^{(s)} X^{m_1} Y^{n_1} Z^{\nu-1} + \cdots + \epsilon_{\nu-1}^{(s)} X^{m_{\nu-1}} Y^{n_{\nu-1}} Z + \epsilon_\nu^{(s)} X^{m_\nu} Y^{n_\nu}.$$

The equations of the transformation from $F^{(s)}$ to F_α are the following [see (53)]:

(58) $$X = X_\alpha^{g_{\alpha-1}} Y_\alpha^{g_\alpha}, \qquad Y = X_\alpha^{f_{\alpha-1}} Y_\alpha^{f_\alpha}, \qquad Z = X_\alpha^p Y_\alpha^q Z_\alpha.$$

Since by the transformation (58) the element ω_s is transformed into Ω_α (after a suitable power of X_α^ν and of Y_α^ν has been deleted), it follows, in view of the inequalities (56), that if we put

(59) $$\begin{aligned} M_j &= m_j g_{\alpha-1} + n_j f_{\alpha-1}, \\ N_j &= m_j g_\alpha + n_j f_\alpha, \end{aligned}$$

then

(60) $$p = \text{minimum } [M_j/j], \qquad q = \text{minimum } [N_j/j],$$

and

(61) $$\bar{M}_j = M_j - pj, \qquad \bar{N}_j = N_j - qj.$$

Using (59) we write M_j and N_j in the following form

$$\begin{aligned} M_j &= g_{\alpha-1}[m_j + n_j\tau + n_j(f_{\alpha-1}/g_{\alpha-1} - \tau)], \\ N_j &= g_\alpha[m_j + n_j\tau + n_j(f_\alpha/g_\alpha - \tau)]. \end{aligned}$$

Since $\left|\dfrac{f_{\alpha-1}}{g_{\alpha-1}} - \tau\right| < \dfrac{1}{g_{\alpha-1}g_\alpha}$ and $\left|\dfrac{f_\alpha}{g_\alpha} - \tau\right| < \dfrac{1}{g_{\alpha-1}g_\alpha}$, it follows that

$$\begin{aligned} &g_{\alpha-1}(m_j + n_j\tau) - M_j \to 0 \\ &g_\alpha(m_j + n_j\tau) - N_j \to 0, \end{aligned}$$

as $\alpha \to \infty$. Hence if α is sufficiently high, then

$$[M_j/j] - 1 \leqq \left[\frac{(m_j + n_j\tau)g_{\alpha-1}}{j}\right] \leqq [M_j/j],$$

$$[N_j/j] - 1 \leqq \left[\frac{(m_j + n_j\tau)g_{\alpha}}{j}\right] \leqq [N_j/j],$$

for $j = 1, 2, \cdots, \nu$ [or better: for all j such that $\epsilon_j^{(s)} \neq 0$]. From these inequalities and from the fact that $g_\alpha \to \infty$, we draw the following consequence: if k is one of the integers $1, 2, \cdots, \nu$ such that

$$\frac{m_k + n_k\tau}{k} \leqq \frac{m_j + n_j\tau}{j}, \quad \text{all } j, \tag{62}$$

then, for α sufficiently high,

$$\begin{cases} [M_j/j] - [M_k/k] \to +\infty, \\ [N_j/j] - [N_k/k] \to +\infty, \end{cases} \tag{63}$$

if in (62) the sign $<$ holds. On the other hand, if the equality sign holds in (62), then $m_k/k = m_j/j$ and $n_k/k = n_j/j$, since τ is irrational, and hence, by (59),

$$\begin{aligned} [M_j/j] &= [M_k/k], \\ [N_j/j] &= [N_k/k]. \end{aligned} \tag{63'}$$

From (63), (63′), (60) and (61) we conclude that if α is sufficiently high, then

$$p = [M_k/k], \qquad q = [N_k/k],$$

while

$$\bar{M}_j \to +\infty, \qquad \bar{N}_j \to \infty$$

if (63) holds, and

$$\bar{M}_j < j, \qquad \bar{N}_j < j,$$

if (63′) holds. In particular, we have

$$\bar{M}^k < k, \qquad \bar{N}^k < k. \tag{64}$$

If minimum $(\bar{M}_j + \bar{N}_j - j) < 0$, then P_α is of multiplicity less than ν for F_α. In the contrary case we apply to F_α a quadratic transformation of center P_α. Interchanging, if necessary, X_α and Y_α, we may assume that the equations of the quadratic transformations are of the form:

$$X_{\alpha 1} = X_\alpha, \qquad Y_{\alpha 1} = Y_\alpha/X_\alpha, \qquad Z_{\alpha 1} = Z_\alpha/X_\alpha.$$

Let $F_{\alpha 1}$ be the transform of the surface F_α and let $P_{\alpha 1}$ be the center of v on $F_{\alpha 1}$. The surface $F_{\alpha 1}$ will be defined in the quotient ring of $P_{\alpha 1}$ by the principal ideal $(\Omega_{\alpha 1})$, where [see (55)]

$$\Omega_{\alpha 1} = \bar{\epsilon}_0^{(\alpha)} Z_{\alpha 1}^{\nu} + \bar{\epsilon}_1^{(\alpha)} X_{\alpha 1}^{\bar{M}_{11}} Y_{\alpha 1}^{\bar{N}_{11}} Z_{\alpha 1}^{\nu-1} + \cdots + \bar{\epsilon}_{\nu-1}^{(\alpha)} X_{\alpha 1}^{\bar{M}_{\nu-1,1}} Y_{\alpha 1}^{\bar{N}_{\nu-1,1}} Z_{\alpha 1} + \bar{\epsilon}_{\nu}^{(\alpha)} X_{\alpha 1}^{\bar{M}_{\nu 1}} Y_{\alpha 1}^{\bar{N}_{\nu 1}},$$

and where

$$\bar{M}_{j1} = \bar{M}_j + \bar{N}_j - j, \qquad \bar{N}_{j1} = N_j,$$

whence, by (64),

$$\bar{M}_{k1} < M_k. \tag{65}$$

If $Z_{\alpha 1} \neq 0$ at $P_{\alpha 1}$, we conclude by the usual argument based on the condition of Lemma 14.1, that $P_{\alpha 1}$ is of multiplicity less than ν for $F_{\alpha 1}$. If $Z_{\alpha 1} = 0$ at $P_{\alpha 1}$, then $X_{\alpha 1}$, $Y_{\alpha 1}$ and $Z_{\alpha 1}$ are uniformizing parameters of $P_{\alpha 1}$. The expression of $\Omega_{\alpha 1}$ is similar to that of Ω_α in (55). Moreover, by (65), we still have $\bar{M}_{k1} < k$, while N_k has not been affected. But since, by (65), the exponent M_k has been replaced by a lower exponent, we conclude that after a finite number of quadratic transformations applied successively to F_α, $F_{\alpha 1}$, $F_{\alpha 2}$, $\cdots$ we must ultimately get a surface F_α on which the center of the valuation v is of multiplicity less than ν. This surface *may* be reached before the sum $\bar{M}_{kl} + \bar{N}_{kl}$ becomes less than k, but at any rate the inequalities

$$\bar{M}_k + \bar{N}_k > \bar{M}_{k1} + \bar{N}_{k1} > \bar{M}_{k2} + \bar{N}_{k2} > \cdots$$

guarantee that it will be reached after a finite number of steps. This concludes the proof of the local reduction theorem.

Part III

Reduction of the singularities of an embedded surface by quadratic and monoidal transformations of the ambient V_3

16. Preparation of the singular locus of the surface F

The singular locus of F—i.e., the set of singular points of F—is a proper algebraic subvariety of F. This statement, whose proof is trivial in the case of ground fields of characteristic zero or of perfect fields of characteristic p, will be proved in all generality in our paper to be printed elsewhere. Anticipating this result, we have therefore that the singular locus of F consists of a finite number of singular curves and of isolated points. A singular curve which is s-fold for F contains at most a finite number of points whose multiplicity for F is greater than s (Lemma 6.5). Hence the multiplicities of the singular points of F have a maximum, say ν, and the singular points of highest multiplicity ν form again a proper algebraic subvariety of F. Let Δ_1, Δ_2, $\cdots$, Δ_h be the irreducible ν-fold curves of F. All points of each Δ_i are exactly ν-fold for F. Each Δ_i may have a finite number of singular points, and these, together with the intersections of pairs of curves Δ_i, constitute the singular locus of the total ν-fold curve $\Delta_1 + \Delta_2 + \cdots + \Delta_h$ of F.

The singularities of each curve Δ_i can be resolved by quadratic transformation of the ambient space V (Theorem 4, section 9). However, a quadratic

transformation will generally introduce new singular curves on the transform of F. We must analyze the situation more closely.

Let T be a quadratic transformation of the ambient V, with center at a singular point P of one of the curves Δ_i, say of Δ_1. Let V' and F' be the T-transforms of V and of F respectively. The quotient ring of any point of F, different from P, is not affected by T, nor is the quotient ring of any irreducible curve on F affected by T. Hence if A' is any point of F' and if A is the corresponding point of F, then $m_{F'}(A') = m_F(A) \leqq \nu$, if $A \neq P$. If $A = P$, then we know (Lemma 3.2) that $m_{F'}(A') \leqq m_F(A)$. Hence ν is still the maximum multiplicity of the singular points of F', and to each curve Δ_i ($i = 1, 2, \cdots, h$) there will correspond on F' an irreducible curve Δ_i' of multiplicity ν. If F' possesses another ν-fold curve Δ', different from each Δ_i', such a curve can only correspond to the point P. But in that case that ν-fold curve Δ' must be irreducible and *free from singularities* (Theorem 1, section 3), and naturally will remain free from singularities under all further successive quadratic transformations which we may have to apply.

We therefore conclude that after a finite number of successive quadratic transformations we shall get a variety V^* and a birational transform F^* of F, on V^*, such that *the irreducible singular curves of F^*, of highest multiplicity ν, are all free from singularities.* We shall assume therefore that the original surface F already satisfies this condition, i.e. we have now that each curve Δ_i is free from singularities, $i = 1, 2, \cdots, h$.

The total ν-fold curve $\Delta_1 + \Delta_2 + \cdots + \Delta_h$ may still have singularities at points which are common to two or more curves Δ_i. To these points we apply quadratic transformations. Let then P be a singular point of the total ν-fold curve of F, and let V' and F' be respectively the transforms of V and of F by the quadratic transformation of center P. If Δ_i goes through P and if $T[\Delta_i] = \Delta_i'$, then Δ_i' will carry exacty one point, say P_i', which corresponds to P, since P is a simple point of Δ_i. If Δ_j is another ν-fold curve through P and if P_j' is the point of Δ_j' ($= T[\Delta_j]$) which corresponds to P, then $P_i' = P_j'$ if and only if the two curves Δ_i and Δ_j have the same tangential direction at P (see section 8). If T creates a new ν-fold curve, say Γ', then, by Lemma 8.4, the tangential direction of Δ_i' at P_i' is different from the tangential direction of Γ' at P_i'. Hence if no two of the curves $\Delta_1', \Delta_2', \cdots, \Delta_h'$ have a common point, then the only singularities of the total ν-fold curve of F' are normal crossings.

In the contrary case we repeat the procedure, i.e. we apply to F' a quadratic transformation T' whose center is a common point of two curves Δ_i' and Δ_j', getting V'', F'' and ν-fold curves $\Delta_1'', \Delta_2'', \cdots, \Delta_h''$. We observe that if T has created a new ν-fold curve Γ', then the ν-fold Γ'' ($= T[\Gamma']$) of F'' will not intersect any of the ν-fold curves $\Delta_1'', \Delta_2'', \cdots, \Delta_h''$, by the remark just made, while its intersection with the ν-fold curve of F'' which possibly has been created by T' will be a normal crossing. We assert that in this fashion we will get, after a finite number of quadratic transformations, models $V^{(s)}$ and $F^{(s)}$ such that no two of the transforms $\Delta_1^{(s)}, \Delta_2^{(s)}, \cdots, \Delta_h^{(s)}$ of $\Delta_1, \Delta_2, \cdots, \Delta_h$ have common

points, and *that consequently the only singularities of the total ν-fold curve of $F^{(s)}$ are normal crossings.* To prove our assertion, let us suppose that the assertion is false. We will have an infinite sequence of models V, F; V', F'; $\cdots$; $V^{(i)}$, $F^{(i)}$; $\cdots$ and an infinite sequence of points P, $P^{(1)}$, $P^{(2)}$, $\cdots$, $P^{(i)}$, $\cdots$ such that: 1) $F^{(i+1)}$ is the transform of $F^{(i)}$ by a quadratic transformation of center $P^{(i)}$; 2) P is a common point of two curves Δ_i, say of Δ_1 and Δ_2, and the successive transforms Δ_1', Δ_1'', $\cdots$; Δ_2', Δ_2'', $\cdots$ are such that $P^{(i)}$ is a common point of $\Delta_1^{(i)}$ and $\Delta_2^{(i)}$. Let Ω denote the union of the quotient rings $Q_{F^{(i)}}(P^{(i)})$. This ring is a proper ring, (i.e. not a field). Hence Ω is contained in the valuation ring of at least one zero-dimensional valuation v. The center of v on $F^{(i)}$ is the point $P^{(i)}$.

We reach the absurd conclusion, in view of Lemma 11.1, that the valuation v must be composed with a divisor whose center on F is at the same time the curve Δ_1 *and* the curve Δ_2. Our assertion is therefore established.

From now on we shall assume that the above preparation of the singular ν-fold locus of F has already been accomplished. Therefore the ν-fold curves $\Delta_1, \Delta_2, \cdots, \Delta_h$ of F are now not only free from singularities, but their mutual intersections are normal crossings and no intersection is common to more than two curves Δ_i. At this stage it may be well to recall our stipulation made in the beginning of section 1, according to which all our considerations apply only to *simple points of the ambient space V.* All points of F, in particular all singular points of F, which fall at singular points of V are excluded.

17. Reduction of the singularities of F (Theorem of Beppo Levi)

In section 10 we have defined permissible transformations, and we have observed that, in general, there is a certain degree of freedom in the selection of a sequence of permissible transformations. This was due to the fact that if a ν-fold point P lies on a ν-fold curve Δ, then a quadratic transformation of center P *and* a monoidal transformation of center Δ are both permissible. We shall now restrict the type of permissible transformations to be used, by stipulating that *a quadratic transformation of center P shall be used only if P is an isolated ν-fold point.* A sequence of permissible transformations in which each transformation is restricted according to the above stipulation shall be called *a normal sequence of permissible transformations.* If then

$$F,\ F_1,\ F_2,\ \cdots,\ F_i,\ \cdots$$

is a sequence of birational transforms of F by successive permissible quadratic and monoidal transformations T, T_1, T_2, $\cdots$, T_i, $\cdots$ which form a normal sequence, then the following conditions are satisfied:

1) either T_i is a monoidal transformation whose center is a ν-fold curve of F_i,

2) or T_i is a quadratic transformation whose center is an isolated ν-fold point of F_i.

It is clear that the theorem of reduction of singularities of the algebraic surface F is established if it can be proved that by a suitable birational transformation

it is possible to lower the maximum of the multiplicities of the singular points of F. This has been established by Beppo Levi [2] in the classical case. We state the theorem of Beppo Levi in the following form:

THEOREM 6. *Let ν be the maximum multiplicity of the singular points of F and let the total ν-fold curve of F have only normal crossings as singularities. Under these conditions, if*

$$F,\ F_1,\ F_2,\ \cdots,\ F_i,\ \cdots \tag{66}$$

is a sequence of surfaces obtained from F by permissible transformations forming a normal sequence (the center of each transformation being a ν-fold point or a ν-fold curve), then the sequence (66) *is necessarily finite.*

Before we proceed with the proof of Theorem 6, we make a few observations concerning normal sequences of permissible transformations which will clarify the reduction process called for by this theorem. In building up a normal sequence of permissible transformations we have a certain degree of freedom, for at each step we may select any ν-fold curve or any isolated ν-fold point as center of the transformation. But granted that each *maximal* normal sequence is finite, it is not difficult to see that *any two maximal normal sequences of transformations lead to one and the same surface.* This can be seen as follows. First, if $P^{(1)}$ and $P^{(2)}$ are two isolated ν-fold points on F, and if we apply to F a quadratic transformation of center $P^{(1)}$, getting a surface F_1, and then apply to F_1 a quadratic transformation whose center is the transform of the point $P^{(2)}$, getting a surface F_2, then it is clear that the same surface F_2 would be obtained if the points $P^{(1)}$, $P^{(2)}$ were interchanged. Similarly, if P and Δ are respectively a ν-fold point and a ν-fold curve of F, then P is a permissible center of a quadratic transformation only if P is isolated, hence at any rate not on Δ. But then it is clear that also in this case a quadratic transformation of center P followed by a monoidal transformation whose center is the transform of Δ, leads to the same surface as does a monoidal transformation of center Δ followed by a quadratic transformation whose center is the transform of P. This possibility of interchanging two consecutive transformations of a normal sequence of transformations extends also to the case where both transformations are monoidal whose centers are represented by ν-fold curves of F. This is obvious if the two ν-fold curves do not meet. If they do meet, then their intersections are normal crossings, and our assertion follows from Lemma 8.3.

It follows that the context of Theorem 6 is not narrowed down to a more special procedure if we, for instance, proceed as follows. If the surface F possesses ν-fold curves, we apply first only monoidal transformations. Then Theorem 6 says, in part, that after a finite number of steps all ν-fold curves will be eliminated[9]. *This, however, we know already from Theorem* 4′ (section 9). To the new surface, free from ν-fold curves, which is thus obtained, we now

[9] We again recall that our considerations are limited only to the points or curves on F which are not singular for the ambient V_3.

apply successive quadratic transformations whose centers are isolated ν-fold points. This second stage of the process terminates whenever a quadratic transformation creates a new ν-fold curve. This curve must then first be resolved by monoidal transformations before quadratic transformations come again into play. Theorem 6 asserts that by this procedure, where quadratic and monoidal transformations alternate according to a well-defined pattern, we must obtain after a finite number of steps a birational transform of F whose singular points are all of multiplicity less than ν.

Let F, F_1, F_2, $\cdots$, be a sequence of birational transforms of F, and let $P, P_1, P_2, \cdots$ be a corresponding sequence of points, $P_i \in F_i$. If T_i denotes the birational transformation from F_i to F_{i+1}, then we shall say that $\{P_i\}$ is a *normal sequence of ν-fold points*, if the following conditions are satisfied: 1) each point P_i is a ν-fold point of its carrier F_i, and is either an isolated ν-fold point or lies on only one ν-fold curve Δ of which it is a simple point, or is a normal crossing of two ν-fold curves; 2) if P_i is an isolated ν-fold point, then T_i is a quadratic transformation with center P_i; 3) if P_i is not isolated, then T_i is a monoidal transformation whose center is a ν-fold curve through P_i; 4) P_i and P_{i+1} are corresponding points under T_i.

It is clear that Theorem 6 is equivalent to the assertion that *a normal sequence of consecutive ν-fold points is necessarily finite*. Let us suppose for a moment that Theorem 6 is false. Then some ν-fold point P of F will give rise to an infinite normal sequence of consecutive ν-fold points

$$(67)\qquad P, P_1, P_2, \cdots, P_i, \cdots.$$

The quotient rings $Q_{F_i}(P_i)$ form a strictly ascending chain of rings, and their union will be contained in the valuation ring of at least one zero-dimensional valuation. Let v be one such valuation. Now *using this valuation v* we apply to F an *arbitrary sequence of permissible transformations* as described in section 10 in connection with the local reduction theorem 5. Let F, F'_1, F'_2, $\cdots$, F'_i, $\cdots$ be the corresponding sequence of consecutive transforms of F, and let P'_i be the center of v on F'_i. By the local reduction theorem there exists *some sequence* of permissible transformations such that only a finite number of the points P'_i are ν-fold. We shall prove, however,—and this will establish Theorem 6—the following lemma:

LEMMA. *If the normal sequence* (67) *of consecutive ν-fold points is infinite, then any sequence of consecutive centers*

$$(68)\qquad P, P'_1, P'_2, \cdots, P'_i, \cdots$$

of the valuation v, obtained by using an arbitrary sequence of permissible transformations, consists entirely of ν-fold points; or, in other words: if a normal sequence of permissible transformations is not capable of lowering the center of a given valuation v (the first center being the point P) *then no sequence of permissible transformations will lower the multiplicity of the center of v.*

This lemma expresses, so to speak, the dominant character of a normal se-

quence of transformations in regard to the reduction process. The proof of this lemma is given in the next section.

18. Proof of the lemma

If two points A and B are members of a *normal* sequence of consecutive ν-fold points, and if A precedes B in that sequence, we shall express this by the notation: $A \prec B$. Thus, assuming—as we do—that (67) is a normal sequence, we have $P_i \prec P_j$ if $i < j$ $(i = 0, 1, 2, \cdots, P_0 = P)$.

We observe that any point P_i in the sequence (67), which is not an isolated ν-fold point, can be followed *immediately* at most by a finite number of ν-fold points which are not isolated. This follows from the fact that any ν-fold curve is ultimately eliminated by monoidal transformations. Let therefore $P_{i_1} = A_1$ and $P_{i_2} = A_2$ be the first two isolated ν-fold points *after* P in the sequence (67). We shall prove that

$$P_1' \prec A_2. \tag{69}$$

The proof of the relation (69) *will establish the lemma.* For the relation (69) implies in the first place that P_1' is a ν-fold point of F_1' (this is part of the definition of the symbol $\prec$). In the second place, it implies that the (uniquely determined) normal sequence of consecutive centers of the valuation v which begins with the ν-fold point P_1' contains the point A_2. Hence that sequence also contains the points $P_{i_2+1}, P_{i_2+2}, \cdots$, since

$$P_{i_2}, P_{i_2+1}, P_{i_2+2}, \cdots$$

is *the* normal sequence of ν-fold points which begins with P_{i_2}. Therefore, under the assumption of the lemma, also the normal sequence of consecutive centers of v which begins with P_1' is infinite. Thus the assumption is true not only for the point P, but also for P_1'. But then, by the same argument, also P_2' is a ν-fold point of F_2', and the normal sequence of consecutive centers of v which begins with P_2' is also infinite, and so on. Consequently all the points P_i' are ν-fold as asserted.

We shall now proceed to prove the relation (69). Let T and T' denote the birational transformation respectively from F to F_1 and from F to F_1'. *If P is an isolated ν-fold point,* then a quadratic transformation of center P is the only permissible transformation. Hence in this case $T = T'$, $F_1 = F_1'$, $P_1' = P_1$, and there is nothing to prove (since $P_1 \prec A_2$). We therefore may assume that P lies on a ν-fold curve Δ. In that case T is necessarily monoidal. As to T', it may be either monoidal or quadratic. If T is monoidal, then its center is either Δ or another ν-fold curve Γ through P (this last case may arise if P is a normal crossing). In the first case $T = T'$, and again there is nothing to prove. In the second case we have essentially the same situation, since we may interchange the order in which the curves Δ and Γ are treated without affecting the *normal* sequence (67) (only the order of the *non isolated* ν-fold points, i.e. the points between P and A_1 and between A_1 and A_2, will be affected).

Hence it is sufficient to prove relation (69) under the following assumptions: *T is a monoidal transformation whose center is a ν-fold curve Δ through P, and T' is a quadratic transformation of center P.*

We select uniformizing parameters t_1, t_2, t_3 of $P(V)$ in such a fashion that t_1 and t_2 are uniformizing parameters of $\Delta(V)$. Let V_1 and V_1' denote respectively the transforms of V under T and T', and let F be defined in $Q(P)$ by the principal ideal (ω). Since P_1 is a ν-fold point, we have [section 7, b)] that T may be assumed to be given by the equations (10) of that section, i.e.:

(70) $$t_{11} = t_1, \qquad t_{12} = t_2/t_1, \qquad t_{13} = t_3,$$

where t_{11}, t_{12} and t_{13} are uniformizing parameters of $P_1(V_1)$; and that ω has the form (see the proof of Lemma 6.3):

(71) $$\omega = t_2^\nu + \sum_{i=0}^{\nu} \alpha_i t_1^i t_2^{\nu-i},$$

where the α_i are non units in $Q_V(P)$. As to the equations of the quadratic transformation T' we must consider separately two cases according as t_3/t_1 is or is not finite at P_1'.

First case. *t_3/t_1 is finite at P_1'.* Let $\tau_3 = t_3/t_1$ and let a be the residue of τ_3 at P_1'. Denoting by k^* the residue field of P, let a be a root of an irreducible polynomial $h^*(u)$ with coefficients in k^* and let $h(\tau_3)$ be a polynomial with coefficients in $Q_V(P)$ such that $h(u)$ reduces to $h^*(u)$ if the coefficients of $h(u)$ are replaced by their P-residues. Then t_1, t_2/t_1, and $h(\tau_3)$ are uniformizing parameters of $P_1'(V_1')$, so that we can write the equations of T' as follows:

(72) $$t_1' = t_1, \qquad t_2' = t_2/t_1 = t_{12}, \qquad t_3' = h(\tau_3) = h(t_3/t_1).$$

The quotient ring $Q_{V_1}(P_1)$ consists of all quotients $f(t_2')/g(t_2')$, where f and g are polynomials with coefficients in $Q_V(P)$ and where $g^*(0) \neq 0$ [see section 5, d)][10].

Similarly [see section 2, c)], the quotient ring $Q_{V_1'}(P_1')$ consists of all quotients $f(t_2', \tau_3)/g(t_2', \tau_3)$, where f and g are polynomials with coefficients in $Q_V(P)$ and where $g^*(0, a) \neq 0$. This shows that

(73) $$Q_{V_1}(P_1) \subseteq Q_{V_1'}(P_1').$$

Both surfaces F_1 and F_1' are defined in the respective quotient rings $Q_{V_1}(P_1)$ and $Q_{V_1'}(P_1')$ by the principal ideal (ω_1), where $\omega_1 = \omega/t_1^\nu$, i.e., by (71),

(74) $$\omega_1 = t_{12}^\nu + \sum_{i=0}^{\nu} \alpha_i t_{12}^{\nu-i}.$$

Since P_1 is a ν-fold point of F_1, we must have $\omega_1 \equiv 0(t_{11}, t_{12}, t_{13})^\nu$ in $Q_{V_1}(P_1)$. Hence $\omega_1 \equiv 0(t_1', t_2', t_1' \cdot \tau_3)^\nu \equiv 0(t_1', t_2')^\nu$. This shows that the curve L' defined

[10] Here and throughout this section, an asterisk affixed to a polynomial with coefficients in $Q_V(P)$ indicates that the coefficients of the polynomial have been replaced by their P-residues.

by the ideal (t_1', t_2') is a ν-fold curve of F_1'. This ν-fold curve is obviously one which has been created by the quadratic transformation T'.

We now apply to V_1' a monoidal transformation M with center L'. Let V_2^* and F_2^* be the transforms of V_1' and of F_1' respectively, under M, and let P_2^* be the center of the valuation v on F_2^*. It has been pointed out above that the elements α_i in (71) are non units in $Q_V(P)$. Hence all α_i belong to the principal ideal (t_1') *in the ring* $Q_{V_1'}(P_1')$. Therefore, by (74), $\omega_1 \equiv t_2'^{\nu}(t_1')$. Since $\omega_1 \equiv 0(t_1', t_2')^{\nu}$, and since $\omega_1 = 0$ on F, it follows immediately that $v(t_2') \geqq v(t_1')$[11], and *consequently* t_2'/t_1' *is finite* at P_2^*. Therefore if we put $\tau^* = t_2'/t_1'$ and if we denote by b the P_2^*-residue of τ^* (b = v-residue of τ^*), then the quotient ring $Q_{V_2^*}(P_2^*)$ consists of all quotients $f'(\tau^*)/g'(\tau^*)$, where f' and g' are polynomials with coefficients in $Q_{V_1'}(P_1')$ and where $g'^*(b) \neq 0$; here g'^* is the polynomial obtained from g' by replacing the coefficients of g' by their P'-residues.

On the other hand, let us apply to F_1 a quadratic transformation of center P_1 and let us denote by V_2^{**} and F_2^{**} the transforms of V_1 and F_1 respectively. Let P_2^{**} be the center of v on F_2^{**}. We know already that the quotients t_{13}/t_{11} ($= t_3/t_1 = \tau_3$) and t_{12}/t_{11} ($= t_2'/t_1' = \tau^*$) have finite residue in the valuation v, hence are finite at P_2^{**}. Let us compare the two quotient rings $\mathfrak{J}^* = Q_{F_2^*}(P_2^*)$ and $\mathfrak{J}^{**} = Q_{F_2^{**}}(P_2^{**})$. We have: $Q_V(P) \subset Q_{V_1}(P_1) \subset \mathfrak{J}^{**}$, and also $t_2' \in \mathfrak{J}^{**}$, $\tau_3 \in \mathfrak{J}^{**}$. From this it follows that $Q_{V_1'}(P_1') \subseteqq \mathfrak{J}^{**}$. Since also τ^* is contained in $\mathfrak{J}^{**}$ we conclude that

$$\mathfrak{J}^* \subseteqq \mathfrak{J}^{**}.$$

On the other hand, since $\tau^* \in \mathfrak{J}^*$ we conclude from (73) that

$$\mathfrak{J}^{**} \subseteqq \mathfrak{J}^*.$$

Therefore $\mathfrak{J}^* = \mathfrak{J}^{**}$, i.e. *the point* P_2^* *can be obtained directly from the point* P_1 *by applying to* F_1 *a quadratic transformation.* Before we draw conclusions from this result, we consider the second case.

SECOND CASE. $t_1/t_3 = 0$ at P_1'. The uniformizing parameters of $Q_{V_1'}(P_1')$ are now:

$$(75) \qquad t_1' = t_1/t_3, \qquad t_2' = t_2/t_3, \qquad t_3' = t_3,$$

and $Q_{V_1'}(P_1')$ consists of all quotients $f(t_1', t_2')/g(t_1', t_2')$ of polynomials in t_1', t_2' with coefficients in $Q_V(P)$, such that $g(0, 0) \neq 0$. The surface F_1' is now defined in $Q_{V_1'}(P_1')$ by the principal ideal (ω_1'), where [see (71)],

$$\omega_1' = \omega/t_3^{\nu} = t_2'^{\nu} + \sum_{i=0}^{\nu} \alpha_i t_1'^{i} t_2'^{\nu-i}.$$

The curve Δ_1' defined by the ideal (t_1', t_2') is ν-fold for F_1'. This curve is merely the transform Δ' of the ν-fold curve Δ under T'.[12] We shall show now that

[11] The elements t_1' and t_2' are now thought of as elements of the field of rational functions on the surface F_1'.

[12] This shows incidentally that the second case under consideration arises when the point P_1' corresponds to the tangential direction of Δ at P.

there is another ν-fold curve of F_1' through P_1', namely the curve L' defined by the ideal (t_2', t_3'). To see this, we prove that ω can be written as a finite sum of the form

$$\omega = \sum_i \beta_{(i)} t_1^{i_1} t_2^{i_2} t_3^{i_3}, \qquad \beta_{(i)} \in Q_V(P), \tag{76}$$

where the exponents i_1, i_2, i_3 in each term satisfy the inequality:

$$i_1 + 2i_2 + i_3 \geqq 2\nu. \tag{77}$$

The proof is immediate. We observe that if we have a monomial $\beta t_1^{i_1} t_2^{i_2} t_3^{i_3}$ in which β is a non-unit in $Q_V(P)$, then $\beta = \beta_1 t_1 + \beta_2 t_2 + \beta_3 t_3$, $\beta_i \in Q_V(P)$, and consequently the monomial can be written as a sum of three similar monomials with higher exponents. Hence we can certainly write ω as a finite sum of the form:

$$\omega = \sum_j \epsilon_{(j)} t_1^{j_1} t_2^{j_2} t_3^{j_3} + \sum_i \beta_{(i)} t_1^{i_1} t_2^{i_2} t_3^{i_3},$$

where the $\epsilon_{(j)}$ are units in $Q_V(P)$, $j_1 + 2j_2 + j_3 < 2\nu$ and $i_1 + 2i_2 + i_3 \geqq 2\nu$. Since Δ is ν-fold for F, we also must have $j_1 + j_2 \geqq \nu$, $i_1 + i_2 \geqq \nu$. We find then [see (74)]:

$$\omega_1 = \sum_j \epsilon_{(j)} t_{11}^{j_1+j_2-\nu} t_{12}^{j_2} t_{13}^{j_3} + \sum_i \beta_{(i)} t_{11}^{i_1+i_2-\nu} t_{12}^{i_2} t_{13}^{i_3}.$$

The point P_1 being ν-fold for F_1, we must have $\omega_1 \equiv 0(t_{11}, t_{12}, t_{13})^\nu$. Each term in the second sum belongs to the ideal $(t_{11}, t_{12}, t_{13})^\nu$, in view of the inequality $i_1 + 2i_2 + i_3 \geqq 2\nu$. On the other hand, no term of the first sum belongs to the ideal $(t_{11}, t_{12}, t_{13})^\nu$, since $j_1 + 2j_2 + j_3 < 2\nu$ and since $\epsilon_{(j)}$ are units in $Q_V(P)$. Since no two terms in the first sum have the same exponents, we have a contradiction, unless the first sum is not present at all. This proves that ω is of the form (76) where the exponents satisfy (77). From (76) we obtain the following expression of ω_1':

$$\omega_1' = \sum_i \beta_{(i)} t_1'^{i_1} t_2'^{i_2} t_3'^{i_1+i_2+i_3-\nu},$$

and since in each term the sum of the exponents of t_2' and t_3' is not less than ν, in view of (77), the curve L' is indeed ν-fold for F_1', as asserted.

The point P_1' is therefore a normal crossing of the two ν-fold curves Δ_1' and L'. We now apply to F_1' a birational transformation M' combined of two monoidal transformations of centers Δ_1' and L' respectively. The order in which these two monoidal transformations are applied is immaterial (see Lemma 8.3). Let V_2^* and F_2^* be the transforms of V_1' and F_1' respectively, and let P_2^* be the center of v on F_2^*. From the expression (74) of ω_1 we find that t_{12}/t_{13} has finite v-residue [i.e. $v(t_{12}/t_{13}) \geqq 0$], since $v(t_{11}) > v(t_{13})$ and we cannot have simultaneously $v(t_{12}) < v(t_{11})$, $v(t_{12}) < v(t_{13})$. Now $t_{12}/t_{13} = t_2/t_1t_3 = t_2'/t_1't_3'$. Hence, taking into

account the equations (14) of section 8, we conclude that the equations of M' are of the form

$$t_1^* = t_1', \qquad \tau_2^* = t_2'/t_1't_3', \qquad t_3^* = t_3',$$

where t_1^*, t_3^* are two of the uniformizing parameters of $P_2^*(V_2^*)$, while τ_2^* is not necessarily zero at P_2^*, but is at any rate finite at P_2^*. From these equations of M' and from (75) and (70) we find

$$t_1^* = t_{11}/t_{13}, \qquad \tau_2^* = t_{12}/t_{13}, \qquad t_3^* = t_{13},$$

and from these equations we conclude as in the preceding case that *the point P_2^* can be obtained directly from the point P_1 by a quadratic transformation of center P_1.*

In either case we have the following situation: a) $P_1' < P_2^*$, since F_2^* is obtained from F_1' by one or two monoidal transformations with center at ν-fold curves; b) P_2^* is obtained from P_1 by a quadratic transformation of center P_1. If $i_1 = 1$, i.e. if P_1 is isolated, then $P_2^* = P_2$ and (69) follows, since $P_1' < P_2^* = P_2$ and either P_2 is A_2 or $P_2 < A_2$. If $i_1 > 1$, then P_1 and P_2^* are in the same relation as P and P_1', and there will therefore exist a point P_3^* such that: a) $P_2^* < P_3^*$; b) P_3^* is obtainable directly from P_2 by a quadratic transformation. If then $i_1 = 2$, then $P_3^* = P_3$, and since $P_1' < P_3^*$ the relation (69) follows. If $i_1 > 2$, we repeat the same procedures. Ultimately, after i_1 steps we will be able to conclude that $P_1' < A_2$. This completes the proof of the lemma and also of Theorem 6.

Part IV

REDUCTION OF THE SINGULARITIES OF THREE-DIMENSIONAL VARIETIES

19. Elimination of the simple fundamental locus of a birational transformation of a three-dimensional variety

Let V and V' be two birationally equivalent three-dimensional irreducible algebraic varieties. We denote by T the birational transformation of V into V'. We mean by the "simple fundamental locus" of T the set of fundamental curves and of isolated fundamental points of T which are *simple* for V.

Suppose that by another birational transformation T^* we have succeeded in transforming V into a variety V^* in such a fashion that the following two conditions are satisfied: a) T^{*-1} has no fundamental points on V^*; b) if a point P^* of V^* corresponds to a simple point of V then P^* is not fundamental for the birational correspondence between V^* and V'. Under these conditions we shall say that *the birational transformation T^* has eliminated the simple fundamental locus of T.*

A preliminary but essential step in our reduction of singularities of three-dimensional varieties, a step to which most of this last part of our investigation has to be dedicated, consists in proving the following theorem:

THEOREM 7. *Given a birational transformation T of a three-dimensional variety V into another variety V', it is possible to eliminate the simple fundamental locus of*

T by successive quadratic and monoidal transformations and to carry out this elimination in such a fashion that simple points of V are transformed into simple points by the successive transformations.

In the proof of this theorem we must therefore show that there exists a sequence of quadratic and monoidal transformations to be applied to V and to the successive transforms of V, such that if T^* denotes the product of these transformations and if V^* denotes the transform of V by T^*, then the following conditions are satisfied:

a) If a point P^* of V^* corresponds to a simple point of V, then P^* is not fundamental for the birational correspondence between V^* and V'.

b) If a point P^* of V^* corresponds to a simple point of V, then P^* itself is a simple point of V^*.

c) T^{*-1} has no fundamental points (on V^*).

Since the only transformations to be used are quadratic and monoidal, not only is condition c) automatically satisfied, but we also have that to every point of V^* there will correspond a unique point of V. Moreover, the quadratic and monoidal transformations which will actually be used will always have their centers on the simple fundamental locus of T and on the successive transforms of that locus. Hence T^* will not affect any point of V which is not fundamental for T. Or, more precisely: if P^* is any point of V^* and if the corresponding point P of V is not fundamental for T, then $Q_V(P) = Q_{V^*}(P^*)$ [whereas for an arbitrary point P^* of V^* we can only assert that $Q_V(P) \subseteq Q_{V^*}(P^*)$].

For the purpose of the proof of Theorem 7 we find it necessary to re-formulate that theorem in terms of *linear systems of surfaces* on V and of the *base loci* of such systems. Quite generally, let Σ be a field of algebraic functions of r independent variables, and let $\omega_0, \omega_1, \cdots, \omega_n$ be elements of some extension field of Σ such that the quotients ω_i/ω_j are elements of Σ. If V is a given model of Σ, the elements ω_i determine a unique set of divisors $\mathfrak{A}_0, \mathfrak{A}_1, \cdots, \mathfrak{A}_n$ of Σ, of the first kind with respect to V, free from common factors and such that

$$\omega_i/\omega_j = \mathfrak{A}_i/\mathfrak{A}_j.$$

Each divisor $\mathfrak{A}_i$ determines on V (and—if V is normal—is defined by) an $(r-1)$-dimensional subvariety F_i whose irreducible components are counted to well-defined multiplicities. If $\lambda_0, \lambda_1, \cdots, \lambda_n$ are arbitrary elements in the ground field k, not all zero, then we can write

$$(\lambda_0\omega_0 + \lambda_1\omega_1 + \cdots + \lambda_n\omega_n)/\omega_0 = \mathfrak{A}(\lambda)/\mathfrak{A}_0,$$

where $\mathfrak{A} = \mathfrak{A}(\lambda)$ is a well-defined divisor of Σ of the first kind with respect to V. Let $F = F(\lambda)$ be the $(r-1)$-dimensional subvariety of V defined by $\mathfrak{A}(\lambda)$. The set $|F|$ of all subvarieties F obtained by letting the λ's vary in k is called a *linear system*; it is the linear system determined by the elements $\omega_0, \omega_1, \cdots, \omega_n$, or rather by the linear (projective) space spanned by the element $\rho(\lambda_0\omega_0 + \lambda_1\omega_1 + \cdots + \lambda_n\omega_n)$, where ρ is an arbitrary factor of proportionality. This system $|F|$ contains of course the particular members $F_0, F_1, \cdots, F_n$, and

since the divisors $\mathfrak{A}_0, \mathfrak{A}_1, \cdots, \mathfrak{A}_n$ have no common factor *the system* $|F|$ *is free from fixed components* (a fixed component being an $(r - 1)$-dimensional variety common to all varieties of the system.)

Let $\eta_0, \eta_1, \cdots, \eta_m$ be the homogeneous coördinates of the general point of V, and let $y_0, y_1, \cdots, y_m$ be the corresponding homogeneous coördinates in the ambient projective space of V. Since the quotients ω_i/ω_j belong to Σ, the elements $\omega_0, \omega_1, \cdots, \omega_n$ are proportional to forms $f_0(\eta), f_1(\eta), \cdots, f_n(\eta)$ of like degree in $\eta_0, \eta_1, \cdots, \eta_m$, with coefficients in the ground field k. It is then clear that the linear system $|F|$ is cut out on V, outside of fixed components, by the linear system of hypersurfaces

$$\lambda_0 f_0(y) + \lambda_1 f_1(y) + \cdots + \lambda_n f_n(y) = 0. \tag{78}$$

If $|F'|$ is the linear system determined by $\{\omega_0, \omega_1, \cdots, \omega_n\}$ on another model V' of Σ, then the two linear systems $|F|$ and $|F'|$ are regarded as corresponding linear systems in the birational correspondence between the two models. This fixes without ambiguity the *law of transformation of linear systems* under birational transformations. By stipulating, as we did, that both systems $|F|$ and $|F'|$ be despoiled of fixed components, we have in reality sidetracked certain difficulties which would unavoidably arise upon a more thorough examination of the situation (in the case of algebraic surfaces see our monograph [4], p. 41). However, for our present purpose the transformation law for linear systems as formulated above is entirely adequate.

Let us suppose now that $\omega_0, \omega_1, \cdots, \omega_n$ are the homogeneous coordinates of the general point of a projective model V' of the field Σ[13]. In this case the linear system $|F'|$ determined on V' by $\{\omega_0, \omega_1, \cdots, \omega_n\}$ is merely the system of hyperplane sections of V', according to (78). On any other projective model V the corresponding linear system $|F|$ is therefore the birational image of the system of hyperplane sections of V'. It is thus seen that given an arbitrary linear system $|F|$ on a variety V, and provided that the quotients $\omega_1/\omega_0, \omega_2/\omega_0, \cdots, \omega_n/\omega_0$ generate the field Σ of rational functions on V, then this linear system determines uniquely a birational transformation of V into another variety V' on which the linear system which corresponds to $|F|$ is the system of hyperplane sections, i.e. uniquely to within a projective transformation of V'.

Going back to Theorem 7, let $|F|$ be the linear system of surfaces on V which is the image of the system of hyperplane sections of V'. It is cut out on V by a linear system of hypersurfaces (78). It is known ([8], p. 528) that if H is a point or a curve on V such that the quotient ring $Q_V(H)$ is integrally closed (i.e. if V is *locally normal* at H; see [8], p. 512) then H is fundamental for the birational transformation T if and only if H belongs to the base locus of the

[13] In order that it should be possible to regard the ω's as homogeneous coordinates of the general point of a projective model of Σ, it is necessary and sufficient to have: $\Sigma = k(\omega_1/\omega_0, \omega_2/\omega_0, \cdots, \omega_n/\omega_0)$.

linear system $| F |$, i.e. if H lies on every F of the system. The condition that $Q_V(H)$ be integrally closed is certainly satisfied if H is a simple point or a simple curve of V. Hence *the simple fundamental locus of T coincides with the base locus of $| F |$*, outside those base curves and isolated base points of $| F |$ which are singular for V.

It is now clear how Theorem 7 can be formulated in terms of the linear system $| F |$ and of the base locus of $| F |$. In re-formulating below this theorem, we drop the condition that $| F |$ be the image of the system of hyperplane sections of a variety V' birationally equivalent to V. Thus the theorem which we are going to state presently is somewhat stronger than Theorem 7.

THEOREM 7'. *Given a linear system $| F |$ (free from fixed components) on a three-dimensional variety V, it is possible to transform V by successive quadratic and monoidal transformations into another variety V^* in such a fashion as to satisfy the following two conditions*:

a) *If $| F^* |$ is the linear system (free from fixed components) on V^* which corresponds to $| F |$, then every base point of $| F^* |$ corresponds necessarily to a singular point of V.*

b) *If a point P^* of V^* corresponds to a simple point of V, then P^* is itself a simple point of V^*.*

20. Linear systems free from singular base points: preparation of the base locus of $| F |$

Let the linear system $| F |$ be cut out on V by the system of hypersurfaces (78), and let $F_0, F_1, \cdots, F_n$ be the particular surfaces in the system which are cut out by the hypersurfaces $f_0 = 0, f_1 = 0, \cdots, f_n = 0$ respectively. Let P be a base point of $| F |$ which is simple for V. We say that P is a *singular base point of the linear system* $| F |$ if P is a singular point of *each* surface F in the system. A *simple base point of* $| F |$ is a base point which is non-singular; such a base point must therefore be simple for at least one surface F in the linear system $| F |$.

LEMMA 20.1. *If a simple point P of V is a simple base point of $| F |$, it is necessarily a simple point of at least one of the $n + 1$ surfaces $F_0, F_1, \cdots, F_n$.*

PROOF. Without loss of generality we may assume that $\eta_0 \neq 0$ at P, where—we recall—$\eta_0, \eta_1, \cdots, \eta_m$ denote the homogeneous coördinates of the general point of V. If we put $\omega_i = f_i(1, \eta_1/\eta_0, \cdots, \eta_m/\eta_0)$, then $\omega_0, \omega_1, \cdots, \omega_n$ are elements of $Q_V(P)$. By the theorem of unique factorization in $Q_V(P)$ let δ be the highest common divisor of $\omega_0, \omega_1, \cdots, \omega_n$ and let $\omega_i = \delta\zeta_i$, $\zeta_i \in Q_V(P)$. It is then clear that each surface $F(\lambda)$ of $| F |$ is defined locally, at P, by the principal ideal $(\lambda_0\zeta_0 + \lambda_1\zeta_1 + \cdots + \lambda_n\zeta_n)$ in $Q_V(P)$, i.e. this principal ideal gives all the components of $F(\lambda)$ which pass through P, and the exponents which occur in the decomposition of the element $\lambda_0\zeta_0 + \lambda_1\zeta_1 + \cdots + \lambda_n\zeta_n$ into prime factors give the multiplicities of the corresponding components of $F(\lambda)$. The point P is simple for $F(\lambda)$ if and only if the leading form of $\lambda_0\zeta_0 + \lambda_1\zeta_1 + \cdots + \lambda_n\zeta_n$ is linear (whence $F(\lambda)$ is necessarily locally irreducible at P). For that, however,

it is necessary that the leading form of at least one of the elements ζ_i be linear, and this proves the lemma.

In this and in the next two sections we prove Theorem 7′ in the special case where *the linear system* $|F|$ *has no singular base points* (at simple points of V; *singular points of* V *are left out entirely from our considerations*). In this case the base locus of $|F|$ consists of simple base curves of $|F|$ and of isolated simple base points. Our first step is to eliminated by quadratic transformations the singularities of the total base curve, other than normal crossings. We prove namely the following lemma:

Lemma 20.2. *By successive quadratic transformations it is possible to transform* V *into a variety* V^* *satisfying condition* b) *of Theorem* 7′ *and such that also the following conditions are satisfied*: 1) *if* $|F^*|$ *denotes the linear system on* V^* *which corresponds to* $|F|$, *then* $|F^*|$ *has no singular base points* (*outside the singular points of* V^*); 2) *each irreducible simple base curve of* $|F^*|$ *is free from singularities* (*outside the singular locus of* V^*); 3) *if two irreducible simple base curves of* $|F^*|$ *have a point* P^* *in common* (P^* — *a simple point of* V^*), *then the tangential directions of the two curves at* P^* *are distinct, and* P^* *does not belong to a third irreducible simple base curve.*

Proof. Let P be a simple point of V which is a singular point of the total base curve of $|F|$. We apply to V a quadratic transformation T of center P and we denote by V' and $|F'|$ the T-transforms of V and of $|F|$ respectively. Let Φ' denote the surface, free from singularities [see section 2, a) and b)], which corresponds to the point P. What is the base locus of $|F'|$? In the first place, it is clear that if H' is a base curve or an isolated base point of $|F'|$, then H' either corresponds to a base variety of $|F|$, *of the same dimension* as H', or H' corresponds to the point P. Hence the base of locus of $|F'|$ consists 1) of the proper transform of the base locus of $|F|$ and possibly 2) of a certain number of new base curves and of isolated base points which lie on the surface Φ'. We are interested in these new base curves on Φ'. By hypothesis, P is a simple base point of $|F|$, hence, by Lemma 20.1, P is a simple point of one of the surfaces $F_0, F_1, \cdots, F_n$. Let, say, P be a simple point of F_0. If $\mathfrak{P}$ denotes the divisor determined by the surface Φ', then ζ_0 is exactly divisible by $\mathfrak{P}$ [see section 2, d) and Lemma 3.1; in the present case ζ_0 plays the role of ω, and ν is equal to 1], while each ζ_i is at least divisible by the first power of $\mathfrak{P}$. Hence if $F_0', F_1', \cdots, F_n'$ denote the members of the system $|F'|$ which correspond respectively to $F_0, F_1, \cdots, F_n$, then we can assert that *the surface* Φ' *is not a component of* F_0'. Hence F_0' is the proper transform of F_0, i.e. $F_0' = T[F_0]$. Therefore the intersection of F_0' with Φ' is *an irreducible curve, free from singularities* (Theorem 1, section 3, where again $\nu = 1$). This is the only curve on Φ' which *may* be a base curve of $|F'|$, and this curve is certainly a simple curve of F_0'. We conclude that our quadratic transformation T may create *at most* one new base curve of the linear system under consideration, and that such a new base curve *is necessarily free from singularities* and is a simple base curve. Moreover all its points are simple base points of $|F'|$, for they are all simple for F_0'. The

rest of the proof is achieved by considerations which are identical with those developed in section 16, Part III. We only note that since the quadratic transformations used have centers at simple points, conditions analogous to condition b) of Theorem 7′ and to condition 1) of Lemma 20.2 are certainly satisfied.

From now on we shall therefore suppose that the base locus of our original system $|F|$ satisfies conditions analogous to conditions 2) and 3) of Lemma 20.2.

21. Linear systems free from singular base points: elimination of the base curves by monoidal transformations

We have proved elsewhere ([7], p. 592), for algebraic surfaces, theorems analogous to theorems 7 and 7′. We therefore know that given a linear system of curves on an algebraic surface (free from fixed components) it is possible to eliminate the base points of the system by successive quadratic transformations. Actually this result has been proved again in the present paper: it is an immediate consequence of Lemma 11.1. Only a change of "dimension terminology" is necessary in order to express this result as a statement concerning linear systems $|F|$ of $(r-1)$-dimensional varieties on a V_r and the $(r-2)$-dimensional base varieties of $|F|$. In particular, the above result applies to the base curves of our linear system of surfaces $|F|$ on the three-dimensional variety V, and it can be used as a stepping stone toward the elimination of these base curves by successive monoidal transformations. However, if $|F|$ was an arbitrary linear system, this extrapolation from base points of a linear system of curves on a surface to base curves of a linear system of surfaces on a three-dimensional variety would have met with a serious difficulty. We proceed to present this difficulty and to show why it disappears in the present case where we deal with a linear system $|F|$ free from singular base points.

Let Δ be an irreducible simple base curve of the system $|F|$. If Σ denotes the field of rational functions on V and if we adjoin to the ground field k an element ξ of Σ *whose Δ-residue is transcendental over k*, then over the new ground field $k(\xi)$ the variety V becomes a "surface," the system $|F|$ becomes a linear system of "curves" on the "surface" $V/k(\xi)$ and Δ becomes a base "point" of $|F|$. Let T be a monoidal transformation of V/k, of center Δ, and let V'/k and $|F'|$ be the T-transforms of V/k and of $|F|$ respectively. The transformation T is a quadratic transformation of $V/k(\xi)$ into $V'/k(\xi)$ with center at the "point" $\Delta/k(\xi)$. If the system $|F'|$ on $V'/k(\xi)$ still possesses base "points" which correspond to the point Δ, these points represent base curves of $|F'|$ on V'/k, *which correspond to the curve* Δ under the monoidal transformation T. To these curves we again apply monoidal transformation and so we continue until the base "point" Δ is eliminated.

All this refers to $k(\xi)$ as ground field. Let us see how our monoidal transformation T actually affects the base locus of $|F|$ over our original ground field k. The transform $T[\Delta]$ of Δ is an irreducible surface Φ'. Since Δ is free from singularities (outside the singular locus of V), all points of Φ' which correspond to

simple points of V are simple both for Φ' and for V' [section 5, a)]. Thus condition analogous to condition b) of Theorem 7' is satisfied, since we know (section 5) that any two-dimensional component of the total transform $T\{\Delta\}$ of Δ, other than Φ', must correspond to points of Δ which are singular for V.

Now let P' be a base point of $| F' |$ which corresponds to a simple point P of V. If P is not on Δ, then it has not been affected by T, and hence P', as well as P, is a simple base point. Suppose that P is on Δ. There exists a surface F, say F_0, in $| F |$, which has at P a simple point. If $\mathfrak{P}$ denotes the divisor defined by the irreducible surface Φ', then we know [section 5, e) and Lemma 6.1] that ω_0 is exactly divisible by $\mathfrak{P}$. Since each element ω_i $(i = 0, 1, \cdots, n)$ is divisible at least by the first power of $\mathfrak{P}$, it follows that Φ' is not a component of F_0', where F_0' denotes that member of the linear system $| F' |$ which corresponds to F_0. Hence F_0' is the proper transform of F_0, i.e. $F_0' = T[F_0]$. Since P is a simple point of F_0 and of Δ, it follows that also P' is a simple point of F_0' (Lemma 6.3). Hence the base points of $| F |$, which (outside the singular locus of V) are by hypothesis simple base points, are transformed into simple base points of $| F' |$.

The points of V' which correspond to singular points of V do not interest us. With the understanding that these points are left out altogether from our consideration, we analyse the possible base curves of $| F' |$.

In general, there could be curves on Φ' which are base curves of $| F' |$ and *which correspond to special points*[14] *of* Δ. Such new base curves, created by the monoidal transformation T, are lost when we pass to the new ground field $k(\xi)$, since ξ is not a transcendental on any such curve. The knowledge that the base "point" $\Delta/k(\xi)$ can be eliminated by "quadratic" transformations would have been of little help to us if in the course of eliminating base "points" over $k(\xi)$ we would find again and again new base curves which correspond to special points of the centers of the successive monoidal transformations. Fortunately, *as a consequence of our assumption that* $| F |$ *has no singular base points, this is not the case.* For suppose that $| F' |$ possesses a base curve Γ' which corresponds to a point P of Δ (P-simple for V). Again assuming that P is a simple point of F_0, it would follow by Lemma 6.2 that $m_{F_0}(P) > m_{F_0}(\Delta)$. This is a contradiction, since $m_{F_0}(P) = m_{F_0}(\Delta) = 1$.

We conclude that any base curve of $| F' |$ which lies on Φ' must correspond to Δ itself, and not to a point of Δ. By Theorem 2, section 6, there can only be at most one such curve, say Δ', since Δ is a simple curve for some surface F in the system $| F |$. This curve Δ' is free from singularities, and (Lemma 8.2) has only normal crossings with other base curves of $| F' |$. If we apply the same treatment to Δ' as we have applied to Δ and if we do that for all the base curves of $| F |$, we see that ultimately we will succeed in eliminating all the base curves

[14] We also may have on Φ' a finite number of isolated base points of $| F' |$ which corresponds to special points of Δ. These points do not interest us in this section, but they have to be taken care of in the next section where we carry out the elimination of isolated base points.

of $|F|$, and that, without violating condition b) of Theorem 7′. We therefore assume from now on that the linear system $|F|$ on our original surface has only isolated simple base points (outside of the singular locus of V).

22. Linear systems free from singular base points: elimination of the isolated base points

Let now P be an isolated simple base point of $|F|$. We apply to V a quadratic transformation T of center P and we denote by V' and $|F'|$ the T-transforms of V and of $|F|$ respectively. We know already from section 20 that T will resolve the base point P *either* into a finite number of isolated simple base points of $|F'|$ *or* into a simple base curve Δ' of $|F'|$. In the first case we continue with quadratic transformations, taking as centers the new isolated base points. In the second case we first eliminate the base curve Δ' by monoidal transformation, as in the preceding section. However, after Δ' has been eliminated, isolated base points (all simple) may remain which correspond to special points of Δ' (see footnote 14). To these we apply again quadratic transformations, and so we continue indefinitely. *We assert that this process must terminate after a finite number of steps.* For suppose that the contrary is true. We will have then an infinite sequence of successive transforms of V:

$$V,\ V^{(1)},\ V^{(2)}, \cdots,\ V^{(i)}, \cdots$$

and of successive simple points $P, P^{(1)}, P^{(2)}, \cdots, P^{(i)}, \cdots$, where $P^{(i)} \in V^{(i)}$, with the following properties:

1. $Q(P) \subset Q(P^{(1)}) \subset Q(P^{(2)}) \subset \cdots$
2. If $|F^{(i)}|$ denotes the linear system on $V^{(i)}$ which corresponds to $|F|$, then $P^{(i)}$ is an isolated base point of $F^{(i)}$.

Let $\bar{V}$ be the variety birationally equivalent to V on which the linear system which corresponds to $|F|$ is the system of hyperplane sections (section 19). Since $P^{(i)}$ is an isolated base point of the linear system $|F^{(i)}|$, it is an isolated fundamental point of the birational transformation which carries $V^{(i)}$ into $\bar{V}$. Since on the other hand $P^{(i)}$ is a simple point of $V^{(i)}$, it follows (see [8], p. 532) that to $P^{(i)}$ there must correspond on $\bar{V}$ a pure two-dimensional variety. Let then $H_i = \{H_{i1}, H_{i2}, \cdots\}$ be the set of irreducible surfaces H_i on $\bar{V}$ which correspond to the point $P^{(i)}$. Since $Q(P_i) \subset Q(P_j)$ if $i < j$, it follows that $H \supseteqq H_1 \supseteqq H_2 \supseteqq \cdots$. Since no set H_i is empty and every H_i is a finite set, it follows that they have an element in common, say G. Then G is an irreducible surface on $\bar{V}$ which corresponds to every point $P^{(i)}$. Let $\mathfrak{P}$ be a divisor of the field Σ whose center on $\bar{V}$ is the surface G. This divisor is then of second kind with respect to each variety $V^{(i)}$, its center on $V^{(i)}$ being the point $P^{(i)}$. If then $\mathfrak{p}^{(i)}$ denotes the ideal of non units in $Q(P^{(i)})$, then $v_{\mathfrak{P}}(\mathfrak{p}^{(i)}) > 0$, where $v_{\mathfrak{P}}(\mathfrak{p}^{(i)})$ denotes the least value assumed by the elements of $\mathfrak{p}^{(i)}$ in the discrete valuation defined by the divisor $\mathfrak{P}$. Now, each $V^{(i)}$ is obtained from $V^{(i-1)}$ by a quadratic transformation of center $P^{(i-1)}$, followed up, perhaps, by a finite number of monoidal transformations. Consequently (see proof of Lemma 9.1)

$$v_{\mathfrak{P}}(\mathfrak{p}^{(1)}) > v_{\mathfrak{P}}(\mathfrak{p}^{(2)}) > \cdots > v_{\mathfrak{P}}(\mathfrak{p}^{(i)}) > \cdots > 0.$$

This yields a contradiction (since each $v_{\mathfrak{P}}(\mathfrak{p}^{(i)})$ is a positive integer) and establishes Theorem 7′ for linear systems free from singular base points.

23. Reduction of the singular base points of a linear system of surfaces

Let now $| F |$ be an arbitrary linear system on V. To complete the proof of Theorem 7′ it is only necessary to show that it is possible to eliminate the singular base points of $| F |$ by quadratic and monoidal transformations, always subject to the additional condition that simple points of V be transformed into simple points. This we proceed to show by using on one hand the reduction theorem for surfaces (Theorem 6, section 17) and on the other hand the theorem of Bertini on the variable singular points of a variety which varies in a linear system. The proof of this theorem of Bertini for abstract varieties is given by us in [9]. For our present application we need some of the concepts which are developed in [9] and the abstract formulation of Bertini's theorem. For simplicity of exposition we shall confine ourselves to three-dimensional varieties.

Denoting as usual the homogeneous coördinates of the general point of V by $\eta_0, \eta_1, \cdots, \eta_m$, let $| F |$ be our linear system (78) on V, free from fixed components, and let $t_0, t_1, \cdots, t_n$ be indeterminates (i.e. we assume that the t's are algebraically independent over the field $k(\eta_0, \eta_1, \cdots, \eta_m)$. We adjoin to the ground field k these indeterminates and we denote by K the extended field $k(t_0, t_1, \cdots, t_n)$. The variety V/k can be regarded as a three-dimensional variety V/K over the new ground field, with the same general point $(\eta_0, \eta_1, \cdots, \eta_m)$ as V.

Every irreducible subvariety W/k of V/k, with general point $(\bar\eta_0, \bar\eta_1, \cdots, \bar\eta_m)$ defines a subvariety W/K of V/K, of the same dimension as W, with the same general point as W and with the property that $t_0, t_1, \cdots, t_n$ are algebraically independent over $k(\bar\eta_0, \bar\eta_1, \cdots, \bar\eta_m)$. The subvariety W/K shall be called *the extension of the subvariety* W/k. Conversely, if an irreducible subvariety W/K of V/K, with general point $(\bar\eta_0, \bar\eta_1, \cdots, \bar\eta_m)$, is such that $t_0, t_1, \cdots, t_n$ are algebraically independent over $k(\bar\eta_0, \bar\eta_1, \cdots, \bar\eta_m)$, then W/K is the extension of a unique subvariety W/k of V/k.

It can be proved that *the singular locus of V/K is the extension of the singular locus of V/k.*

For an *arbitrary* irreducible subvariety W^*/K of V/K, with general point $(\overset{*}{\eta}_0, \overset{*}{\eta}_1, \cdots, \overset{*}{\eta}_m)$, it is true that there exists a unique subvariety W/k of V/k, with general point $(\bar\eta_0, \bar\eta_1, \cdots, \bar\eta_m)$, such that

$$k[\overset{*}{\eta}_0, \overset{*}{\eta}_1, \cdots, \overset{*}{\eta}_m] \cong k(\bar\eta_0, \bar\eta_1, \cdots, \bar\eta_m),$$

where the isomorphism is such that $\overset{*}{\eta}_i$ and $\bar\eta_i$ are corresponding elements. The variety W/k shall be called the *contraction* of W^*/K. We have always:

$$W^*/K \subseteq \text{extension } W/K \text{ of } W/k,$$

whence dimension of $W^*/K \leqq$ dimension of W/k.

The essence of our formulation of Bertini's theorem consists in regarding the *linear system* of surfaces (78) on V/k as *one irreducible surface* F^*/K on V/K. This surface is defined by a general point $(\zeta_0, \zeta_1, \cdots, \zeta_m)$ satisfying the following conditions:

1) $k[\zeta_0, \zeta_1, \cdots, \zeta_m] \cong k[\eta_0, \eta_1, \cdots, \eta_m]$,
whence the contraction of F^*/K is the entire variety V/k.

2) $t_0f_0(\zeta) + t_1f_1(\zeta) + \cdots + t_nf_n(\zeta) = 0$.
It is proved that *if a variety W/K is an extension of W/k then W/K lies on F^*/K if and only if W/k belongs to the base locus of the linear system $|F|$, and that W/K is singular for F^*/K if and only if W/k is a singular base variety of $|F|$.*

After these preliminaries we state the theorem of Bertini on the variable singular points of $|F|$ as follows:

Theorem of Bertini. *If W_1/K is an irreducible singular subvariety of F^*/K, then the contracted subvariety W/k of V/k is either singular for V/k or belongs to the base locus of the linear system $|F|$.*

From this theorem and on the basis of the preliminary results stated above, we conclude that, outside of varieties W_1/K which correspond to singular subvarieties of V/k, the surface F^*/K may possess only the following singular curves and points:

a) Singular curves which are extensions of singular base curves of $|F|$.

b) Isolated singular points which are extensions of isolated singular base points of $|F|$.

c) Isolated singular points whose contractions are simple base curves of $|F|$ (it can be proved that along these curves the surfaces of the linear system have *variable* singular points, *provided the ground field k is of characteristic zero;* see [9].)

If the variety V/k is subjected to a birational transformation which carries V and $|F|$ into V' and $|F'|$ respectively, then V/K and F^*/K undergo a corresponding birational transformation, and the transform F'^*/K of F^*/K has the same relationship to $|F'|$ as F^*/K has to $|F|$. In other words, the surface F^*/K is a birational covariant of the linear system $|F|$. In particular, if we apply to V/k a quadratic or a monoidal transformation with center W/k, then V/K and F^*/K undergo respectively a quadratic or a monoidal transformation whose center is the extension W/K of W/k. If we then apply the reduction Theorem 6 of section 17 to the surface F^*/K and to its ambient variety V/K, we see that by successive monoidal and quadratic transformation of V/k it is possible to resolve the singular curves a) and the singular points b) of the surface F^*/K, i.e. it is possible to resolve the singular base locus of the linear system $|F|$. It should be pointed out that a preparation of the singular curves a) of F^*/K is a preliminary step of the reduction process (section 16). The quadratic transformations involved in this step are actually quadratic transformations over k (i.e. of V/k) since the singular points of the singular curves a) are necessarily extensions of the singular points of the corresponding singular base curves of $|F|$. Finally, we also point out that we do nothing to the isolated singular points c) of F^*/K; they do not correspond to singular base loci of $|F|$ and

their reduction is not necessary for our purposes. Besides, since these singular points are not extensions of points over k, the transformations needed for their reduction are not directly expressable as transformations of V/k.

This completes the proof of Theorems 7 and 7′.

24. A lemma on simple points of an algebraic surface

Let F and F' be two birationally equivalent algebraic surfaces.

LEMMA. *If P and P' are corresponding simple points in the birational correspondence between F and F' and if $Q(P) \subseteq Q(P')$, then either $Q(P) = Q(P')$ or P' can be obtained from P by successive quadratic transformations.*

This lemma implies the following corresponding result for $(r-2)$-dimensional subvarieties of a V_r:

If W and W' are corresponding $(r-2)$-dimensional varieties of a birational correspondence between two r-dimensional varieties V_r and V'_r and if $Q(W) \subseteq Q(W')$, then either $Q(W) = Q(W')$ or W' can be obtained from W by successive monoidal transformations.

PROOF OF THE LEMMA. Assuming that $Q(P) \neq Q(P')$, we have to show that there exists a finite sequence of birational transforms of F:

$$F, F_1, F_2, \cdots, F_n,$$

and a corresponding sequence of points $P_1, \cdots, P_n$, $P_i \in F_i$, such that: 1) F_i is the transform of F_{i-1} by a quadratic transformation of center $P_{i-1}(F_0 = F, P_0 = P)$; 2) P_{i-1} and P_i are corresponding points under this quadratic transformation; 3) $Q(P_n) = Q(P')$.

Let x and y be uniformizing parameters of $P(F)$, and let P_1 be a point of F_1 which corresponds to P' (in the birational correspondence between F_1 and F'). Since $Q(P) \subset Q(P')$, every valuation of center P' on F' has its center at P on V. Consequently also P_1 and P are corresponding points. Without loss of generality we may assume that y/x is finite at P_1. Then x can be taken as one of a pair of uniformizing parameters of $Q_{F_1}(P_1)$.

We shall show now that necessarily $Q(P_1) \subseteq Q(P')$. Assume the contrary. Then P' is a fundamental point of the birational correspondence between F' and F_1. Hence to P' there must correspond on F_1 at least one irreducible curve through P_1. Any such curve must also correspond to P, since $Q(P) \subset Q(P')$. But the irreducible curve Γ_1 defined in $Q(P_1)$ by the principal ideal (x) is the only curve on F_1 which corresponds to P. Hence this is the only curve through P_1 which corresponds to P'.

Let $\mathfrak{P}$ be the divisor defined by the curve Γ_1, and let $\mathfrak{m}$ and $\mathfrak{m}'$ be the ideals of non units in $Q_F(P)$ and $Q_{F'}(P')$ respectively. We have $v_{\mathfrak{P}}(x) = 1$ and $v_{\mathfrak{P}}(m') > 0$. Hence the element x, which is certainly an element of $\mathfrak{m}'$, does not belong to $\mathfrak{m}'^2$. Now P is fundamental for the birational correspondence between F and F', and therefore there must exist on F' an irreducible curve Δ' through P' which corresponds to P. Since both x and y vanish on Δ', this curve must be defined by a prime element of $Q(P')$ which divides in $Q(P')$ both x and y. But x itself is a prime element of $Q(P')$, since $x \not\equiv 0(\mathfrak{m}'^2)$. Hence x

divides y, $y/x \in Q(P')$, whence $Q(P_1) \subseteq Q(P')$. This shows that our assumption that $Q(P_1) \not\subseteq Q(P')$ is absurd.

If $Q(P_1) = Q(P')$, there is nothing more to prove. If $Q(P_1)$ is a proper subring of $Q(P')$, we pass to the quadratic transform F_2 of F_1 taking P_1 as center, and we find as before that F_2 carries a point P_2 such that $Q(P_1) \subset Q(P_2) \subseteq Q(P')$. Since the union of the quotient rings of successive corresponding points P, P_1, $P_2, \cdots$, obtained by successive quadratic transformations, is a valuation ring (see [5], p. 681, Theorem 10; also [7], p. 591, Lemma 2, footnote 14) and since $Q(P')$ is not a valuation ring nor can it contain a valuation ring, it follows that after a finite number of quadratic transformations we must get a point P_n such that $Q(P_n) = Q(P')$, q.e.d.

25. Reduction of the singularities of three-dimensional varieties

Let Σ be a field of algebraic functions of three independent variables over an arbitrary ground field k, and let N be an arbitrary set of zero-dimensional valuations of Σ. A finite set of models $V_1, V_2, \cdots, V_n$ of Σ shall be called a *resolving system* of N if every valuation in N has a simple center on at least one of the modes V_i. The existence of resolving systems for the set of all valuations of Σ will be established by us in a separate paper, for any field Σ of algebraic functions for which the theorem of local uniformization holds true (hence, in particular, for function fields over ground fields of characteristic zero; see [6])[15]. Assuming the existence of a resolving system for the totality of valuations of Σ, we have shown in [7], p. 584 that the theorem of the reduction of singularities is equivalent to the following

FUNDAMENTAL THEOREM. *If there exists a resolving system of N consisting of two models V and V', then there also exists a resolving model for N (i.e. a model of Σ on which every valuation in N has a simple center).*

Our proof of the fundamental theorem will be based exclusively on Theorem 7 of section 19 and on the lemma of the preceding section.

Let V and V' be a resolving pair of N, and let T denote the birational transformation which carries V into V'. We first apply to V the birational transformation T^* of Theorem 7 which eliminates the simple fundamental locus of T on V. Let V^* denote the T^*-transform of V. Since no singular point of V^* corresponds to a simple point of V, also V^* and V' form a resolving pair of N. Let V_1 denote the join of the two varieties V^* and V' (see [8], p. 516). *I assert that also V_1 and V' form a resolving pair for N.* For let v be any valuation in N and let P, P', P^* and P_1 be the centers of v respectively on V, V', V^*, V_1. If P' is a simple point, there is nothing to prove. Let P' be a singular point. Then P is necessarily a simple point. If P is not a fundamental point of T, then $Q(P) \supseteq Q(P')$, and on the other hand $Q(P) = Q(P^*)$, since in this case P has not been affected by the quadratic and monoidal transformations which compose the transformation T^* (see section 19). Hence $Q(P^*) \supseteq Q(P')$ and

[15] For the proof of the existence of finite resolving systems in the classical case (k = field of complex numbers) see [6], p. 855, and in the case of abstract surfaces see [7], p. 592. For the general case see the Bulletin note quoted in footnote **, Introduction.

hence ([8], p. 516) $Q(P_1) = Q(P^*) = Q(P)$. Therefore P_1 is a simple point. If, however, P *is* fundamental for T, then it belongs to the ***simple*** fundamental locus of T, and therefore P^* is not fundamental for the birational transformation between V^* and V'. Hence $Q(P') \subsetneqq Q(P^*) = Q(P_1)$, and since P^* is a simple point of V^* (for P is simple for V) we again conclude that P_1 is a simple point. Thus, one of the two points P_1, P' must be simple, and therefore V_1 and V' constitute indeed a resolving pair of N.

The net effect of the preceding considerations is that we have replaced the original resolving pair (V, V') by the resolving pair (V_1, V'), and for this last resolving pair it is true that the birational correspondence between the two models has no fundamental points on one of them (namely on V_1, since V_1 is the join of V' and of another model). We may therefore assume that this condition is already satisfied for our original resolving pair. *Accordingly, we assume that T^{-1} has no fundamental points on V'.* Then if H and H' are any two corresponding loci of V and V', the following relation will always hold true:

$$Q(H) \subsetneqq Q(H'). \tag{79}$$

We now again apply Theorem 7, and we eliminate the simple fundamental locus of T on V. Let V^* be the birational transform of V which is obtained as a result of this elimination. Again we see that V^* and V' constitute a resolving pair of N.

Let v be an arbitrary valuation in N and let P, P' and P^* be the centers of v on V, V' and V^* respectively. We analyse the relationship between the quotient rings of these three points and we divide this analysis into several cases. For simplicity we denote by F_s the simple fundamental locus of T on V. *We include in F_s all the points of that locus*, hence also points of a simple base curve which are singular for V.

CASE A: $P \notin F_s$. In this case P is either a singular point or is not fundamental for T. In both cases the following is true:

$$\text{(A)} \qquad P' \textit{ is simple and } Q(P') \supseteqq Q(P^*).$$

For $P \notin F_s$ implies that P has not been affected by the birational transformation from V to V^*. Hence $Q(P) = Q(P^*)$, and therefore $Q(P') \supseteqq Q(P^*)$, by (79). Moreover P' must be a simple point even if P is simple, for if P is simple it is not fundamental for T ($P \notin F_s$), i.e. $Q(P) \supseteqq Q(P')$, and hence $Q(P) = Q(P')$, by (79). Hence assertion (A) is established.

CASE B. *P is a simple point and $P \in F_s$.* In this case it is obvious that

$$\text{(B)} \qquad P^* \textit{ is simple and } Q(P^*) \supseteqq Q(P').$$

CASE C. *P is a singular point and $P \in F_s$.* In this case P' is necessarily a simple point. If P' is not fundamental for the birational correspondence between V' and V^*, then we will have:

$$\text{(C}'\text{)} \qquad P' \textit{ is simple and } Q(P') \supseteqq Q(P^*).$$

In (A), (B) and (C′) we have always the same situation, namely of the two points P' and P^* one has the property that it is a simple point and that its

quotient ring contains the quotient ring of the other point. If we now pass to the join of V' and V^*, the center of the valuation v will therefore be a simple point. There remains to consider the case C *under the hypothesis that P' is fundamental for the birational correspondence between V' and V^**. We proceed to study this case.

The point P' is either an isolated fundamental point of the birational transformation between V' and V^* or lies on some fundamental curve of this transformation. We are interested especially in this second possibility. Let Γ' be a *fundamental curve* of the birational transformation between V' and V^* which passes through P'. Since T^{-1} has no fundamental points on V', $T^{-1}(\Gamma')$ is either a unique point or a unique curve on V (note that since P' is simple for V', also Γ' is simple). I *assert that $T^{-1}(\Gamma')$ is necessarily a point.* For suppose the contrary, and let $T^{-1}(\Gamma') = \Gamma$, an irreducible curve on V.

Suppose that $\Gamma \notin F_s$. Then Γ is regular for the birational correspondence between V and V^*. If then Γ^* is the corresponding curve on V^*, then we will have $Q(\Gamma) = Q(\Gamma^*)$. On the other hand we have $Q(\Gamma') \supseteqq Q(\Gamma)$ [by (79)]. Hence $Q(\Gamma') \supseteqq Q(\Gamma^*)$, in contradiction with our assumption that Γ' is fundamental for the birational correspondence between V' and V^*.

Hence Γ must belong to F_s, and is therefore a simple curve of V. We have now the following situation: Γ and Γ' are corresponding *simple* curves of V and V' and, by (79), $Q(\Gamma') \supset Q(\Gamma)$ [not $Q(\Gamma') = Q(\Gamma)$, since $\Gamma \in F_s$]. By the lemma of the preceding section, Γ' can be obtained from Γ by a sequence of monoidal transformations. But in eliminating the simple fundamental curves of T we have used just monoidal transformations. Hence the variety V^* must contain a curve Γ^* such that $Q(\Gamma^*) = Q(\Gamma')$. This is in contradiction with our assumption that Γ' is fundamental for the birational correspondence between V' and V^*.

We have therefore proved that $T^{-1}(\Gamma')$ is necessarily a point. This point is necessarily the point P, since $P' \in \Gamma'$.

Accordingly we consider on V' those isolated fundamental points O' and those fundamental curves Γ' of the birational transformation between V' and V^* *which correspond to singular points of V which lie on F_s and which actually carry centers of valuations in N.* Since (V, V') is a resolving pair of N it is clear that these points O' and these curves Γ' must be simple for V', *and that no point of a Γ' which is singular for V' can occur as center of a valuation in N.* According to Theorem 7 applied to V' and V^* we can eliminate these fundamental points and curves, and by the *italicized* remark just made we do not have to be concerned about the fate of the points of a curve Γ' which are singular for V'. We denote by Ω' the set of the points O' and curves Γ' and we also denote by V_1' the birational transform of V' obtained in the course of the elimination of the fundamental points O' and the fundamental curves Γ'.

Now let again v be any valuation in N and let P, P', P^* and P_1' be the centers of v on V, V', V^* and V_1' respectively. In view of (A), (B), (C′) we find immediately the following:

(A₁) *If $P \notin F_s$, then P_1' is simple and $Q(P_1') \supseteqq (P^*)$.*

For by (A), P' is not fundamental for the birational transformation between V' and V^*. Hence P' is not affected when we pass from V' to V_1', and consequently $Q(P') = Q(P_1')$.

(B$_1$) *If P is simple and $P \in F_s$, then P^* is simple and* $Q(P^*) \supseteqq Q(P_1')$.

For in this case $P' \notin \Omega'$, whence $Q(P') = Q(P_1')$.

The relation (C') yields (since $P' \notin \Omega'$):

(C$_1'$) P_1' *is simple and* $Q(P_1') \supseteqq Q(P^*)$.

There remains the case C, under the hypothesis that $P' \in \Omega'$. In this case, in view of our elimination of the fundamental locus Ω', we must have $Q(P_1') \supseteqq Q(P^*)$, and moreover P_1' is simple since P' is simple. Hence in this case the statement C$_1'$ still holds true.

We have now in all cases, i.e. for any valuation v in N, the situation whereby of the two centers of v on V^* and V_1' one always has the property that it is a simple point and that its quotient ring contains the quotient ring of the second center. *Therefore the join of V^* and V_1' is a resolving model of N.* This completes the proof of the fundamental theorem.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES

[1] W. KRULL, *Dimensionstheorie in Stellenringen*, J. Reine Angew. Math., vol. 179 (1938), pp. 204–226.

[2] BEPPO LEVI: *Risoluzione delle singolarità puntuali delle superficie algebriche*, Atti Accad. Sci. Torino, vol. 33 (1897), pp. 66–86.

[3] F. K. SCHMIDT: *Ueber die Erhaltung der Kettensätze der Idealtheorie bei beliebigen endlichen Körpererweiterungen*, Math. Zeit., vol. 41 (1936), pp. 443–450.

[4] O. ZARISKI: *Algebraic surfaces*, Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 3 (1935).

[5] O. ZARISKI: *The reduction of the singularities of an algebraic surface*, Ann. of Math., vol. 40 (1939), pp. 639–689.

[6] O. ZARISKI: *Local uniformization on algebraic varieties*, Ann. of Math., vol. 41 (1940), pp. 852–896.

[7] O. ZARISKI: *A simplified proof for the resolution of singularities of an algebraic surface*, Ann. of Math., vol. 43 (1942), pp. 583–593.

[8] O. ZARISKI: *Foundations of a general theory of birational correspondences*, Trans. Amer. Math. Soc., vol. 53 (1943), pp. 490–542.

[9] O. ZARISKI: *The theorem of Bertini on the variable singular points of a linear system of varieties*, Trans. Amer. Math. Soc., vol. 56 (1944), pp. 130–140.

ANNALS OF MATHEMATICS
Vol. 45, No. 3, July, 1944

INSTABILITY FOR DOUBLE INTEGRAL PROBLEMS IN THE CALCULUS OF VARIATIONS[1]

BY MAX SHIFFMAN

(Received March 15, 1943)

CONTENTS

I. PRELIMINARIES

1. Introduction

It is the purpose of this investigation to discuss the existence of unstable extremal surfaces bounded by a given rectifiable Jordan curve Γ for a class of double integral problems in parametric form. The double integral problems considered are described in §2 below. The main limitation is that the integral be dominantly an area integral. There are certain indications that the restrictions can be removed, and in the remarks at the end of this paper we shall list all those places where the restrictions were used in an essential way.

The general theory of unstable critical points has had a remarkable development in the theory of M. Morse ([14], [15], [16]).[2] Morse has developed his well

[1] Presented to the American Math. Soc., An abstract appears in the Bull. Amer. Math. Soc., 48, p. 830 (1942).

[2] Numbers in brackets refer to the bibliography at the end.

known relations for the various types of critical points first for functions of a finite number of variables, then for single integral problems in the calculus of variations, and lastly in an abstract way. The abstract theory of Morse lays the topological foundations of a theory of critical points, and analyzes in a general way those analytical steps which remain to be proved in any particular case. The theory thus serves as a valuable guide, but in any application deep analytical considerations are still required.

The methods used by Morse for functions of a finite number of variables and for single integral problems can not be generalized to multiple integral problems. It is well known that multiple integral problems in the calculus of variations have difficulties of a much more serious nature than single integrals. It suffices to note that only in recent years has a theory of the absolute minimum for multiple integrals been developed, and that the theory is far from being in its final form. It is beset with difficulties of a real variable nature. In this paper we shall make use of the fundamental work of L. Tonelli concerning convergence, equicontinuity, and lower semicontinuity for multiple integrals in non-parametric form ([1], [2], [5]), and of the fundamental work of E. J. McShane on integrals in parametric form ([3], [4], [6]).

Concerning instability, it is easy to locate the major difference between multiple integrals and single integrals. For single integrals one can use ordinary differential equation theory to establish the existence and minimizing property of extremals in the small, and decompose a problem in the large into a succession of problems in the small. For multiple integrals this cannot be done—size has no effect.

Instability has been discussed in the literature for one multiple integral problem, namely the Plateau problem. The theory of unstable minimal surfaces has been developed by the author, by M. Morse and C. Tompkins, and by R. Courant ([17], [18], [19]). The most general result as yet in this field has been recently obtained by the author in a brief note [22]. In this note an attempt was made to produce methods which would not be limited solely to the Plateau problem. It is this note which has been generalized here and which serves as a resumé of the ideas involved.

The decisive condition (besides lower semicontinuity, etc.) which Morse requires in order to apply his abstract theory to any particular case is '*reducibility*.' We avoid the necessity of investigating reducibility by imbedding the space into a certain set of larger spaces approaching the original space as a limit. The Morse theory in these larger spaces is readily verified since it depends essentially on the analysis of a *continuous* functional. An additional continuity theorem allows the passage to the limit.

But the passage to the limit necessitates certain topological modifications of the Morse theory (cf. [22]). Accordingly, we shall confine ourselves in the present paper to proving the following theorem: if Γ bounds two extremal surfaces which are proper relative minima, it must bound an unstable extremal surface.

2. Statement of the problem

Let Γ be a closed rectifiable Jordan curve in space. We shall be interested in finding those surfaces bounded by Γ which are extremal surfaces for a double integral of the following form:

$$J[\mathfrak{x}] = \iint \{f(X, Y, Z) + k\sqrt{X^2 + Y^2 + Z^2}\}\, du\, dv$$

where k is a positive constant and X, Y, Z are the Jacobians

$$\begin{vmatrix} y_u & z_u \\ y_v & z_v \end{vmatrix}, \quad \begin{vmatrix} z_u & x_u \\ z_v & x_v \end{vmatrix}, \quad \begin{vmatrix} x_u & y_u \\ x_v & y_v \end{vmatrix}$$

respectively. Surfaces will be written in the vector form $\mathfrak{x} = \mathfrak{x}(u, v)$, with components $x(u, v)$, $y(u, v)$, $z(u, v)$, and

$$\mathfrak{X} = \mathfrak{x}_u \times \mathfrak{x}_v$$

where $\cdot \times \cdot$ is the vector product. The components of $\mathfrak{X}$ are exactly the Jacobians X, Y, Z written above, and $\mathfrak{X}$ is a vector normal to the surface $\mathfrak{x}$. The function $f(X, Y, Z)$ will be written $f(\mathfrak{X})$, and $f_{\mathfrak{X}}$ will mean the vector with components $f_X(X, Y, Z)$, $f_Y(X, Y, Z)$, $f_Z(X, Y, Z)$. The domain of representation in the (u, v)-plane will be the unit circle.

We shall make the following assumptions concerning the function $f(\mathfrak{X})$:

2.1. $f(\mathfrak{X})$ has continuous first derivatives for $\mathfrak{X}^2 = 1$ and is positively homogeneous of degree 1 in $\mathfrak{X}$, i.e., $f(t\mathfrak{X}) = tf(\mathfrak{X})$ if $t \geqq 0$.

2.2. $E(\mathfrak{X}; \mathfrak{X}_0) \geqq 0$, where $E(\mathfrak{X}; \mathfrak{X}_0) = f(\mathfrak{X}) - \mathfrak{X}\cdot f_{\mathfrak{X}_0} = f(\mathfrak{X}) - f(\mathfrak{X}_0) - (\mathfrak{X} - \mathfrak{X}_0) f_{\mathfrak{X}_0}$.

2.3. The minimum m of $f(\mathfrak{X})$ on $\mathfrak{X}^2 = 1$ is > 0.

2.4. The maximum k' of $f(\mathfrak{X})$ on $\mathfrak{X}^2 = 1$ is $< k$.

The main restrictions that are imposed on the integrand are: the integrand does not involve x, y, z explicitly, and the very unnatural condition 2.4 which asserts that J is dominantly an area integral. Conditions 2.1 and 2.2 are usual. Condition 2.3 is unnecessary and can be eliminated, as is well known.

Of all the elementary consequences of these conditions, we shall note merely the following:

$$m\,|\,\mathfrak{X}\,| \leqq f(\mathfrak{X}) \leqq k'\,|\,\mathfrak{X}\,| \tag{2a}$$

$$|\,f_{\mathfrak{X}}\,| \leqq k' \tag{2b}$$

$$|\,f(\mathfrak{X}') - f(\mathfrak{X})\,| \leqq k'\,|\,\mathfrak{X}' - \mathfrak{X}\,| \tag{2c}$$

$$-k'D[\mathfrak{x}] \leqq -k'A[\mathfrak{x}] \leqq \iint \mathfrak{X}\cdot f_{\mathfrak{X}_0}\, du\, dv \leqq k'A[\mathfrak{x}] \leqq k'D[\mathfrak{x}], \tag{2d}$$

where $A[\mathfrak{x}]$ and $D[\mathfrak{x}]$ are defined directly below.

The first decisive step is to replace $J[\mathfrak{x}]$ by the integral

$$I[\mathfrak{x}] = \iint f(\mathfrak{X})\, du\, dv + kD[\mathfrak{x}],$$

where $D[\mathfrak{x}]$ is the Dirichlet-Douglas integral,

$$D[\mathfrak{x}] = \tfrac{1}{2} \iint (\mathfrak{x}_u^2 + \mathfrak{x}_v^2)\, du\, dv = \tfrac{1}{2} \iint (E + G)\, du\, dv,$$

and E, F, G are the first fundamental coefficients,

$$E = \mathfrak{x}_u^2, \qquad F = \mathfrak{x}_u \mathfrak{x}_v, \qquad G = \mathfrak{x}_v^2.$$

This is the starting point of J. Douglas, of T. Radó, and of R. Courant in the Plateau problem, which is the case $f(\mathfrak{X}) = 0$ ([7], [8], [12]). We shall not dwell on the advantages of this replacement; it has been discussed in some detail elsewhere ([9], [10], [12]). Rather, we merely note that the integral $I[\mathfrak{x}]$ depends not only on the surface but on the representation of the surface as well. Out of all possible representations of a surface only those are considered which have a finite Dirichlet integral. And as in the Plateau problem, the extremals for $I[\mathfrak{x}]$ will turn out to be extremal surfaces for $J[\mathfrak{x}]$ given in *isometric* representation.

In addition to the integrals $J[\mathfrak{x}]$, $I[\mathfrak{x}]$, $D[\mathfrak{x}]$, it will be convenient to use the following integrals:

$$F[\mathfrak{x}] = \iint f(\mathfrak{X})\, du\, dv$$

$$A[\mathfrak{x}] = \iint |\, \mathfrak{X} \,|\, du\, dv = \iint \sqrt{EG - F^2}\, du\, dv.$$

If we wish to specify a subset G of the (u, v) unit circle over which the integrals are to be taken, the integrals will be written with a subscript G. The symbol S will always be used to represent a closed interior subdomain of the unit circle. The interior of the circle of radius ρ and center the origin will be indicated by C_ρ, its circumference by C_ρ^*. The annular ring between C_ρ^* and $C_{\rho'}^*$ will be indicated by $C_{\rho\rho'}$. If G is any region of the unit circle, the boundary of G will be designated by G^*.

3. Surfaces of class $\mathfrak{T}$

The surfaces to be admitted to our variational problem must satisfy the following conditions:

3.1. $\mathfrak{x}(u, v)$ is continuous on the closed unit circle.

3.2. $\mathfrak{x}(u, v)$ as a function of u is absolutely continuous for almost all v, and as a function of v is absolutely continuous for almost all u.

3.3. $D[\mathfrak{x}]$ is finite

Surfaces satisfying 3.1 to 3.3 are said to be of class $\mathfrak{T}$.[3] The conditions 3.1 to

---

[3] The class $\mathfrak{T}$ is only one of various possibilities for a class of admitted surfaces. See [5], and K. Friedrichs, *On differential operators in Hilbert spaces*, Amer. Jour. Math., 61, 523–544 (1939).

3.3 are equivalent to requiring $\mathfrak{x}(u, v)$ to be absolutely continuous in the sense of Tonelli and to have a finite Dirichlet integral. All surfaces involved in this paper satisfy 3.1 to 3.3, and so the description "of class $\mathfrak{T}$" will generally speaking be omitted. The known properties of functions absolutely continuous in the Tonelli sense will be used without comment.

For a surface of class $\mathfrak{T}$ it is easily seen that all the integrals A, F, I, J exist. Use will be made of the theorems of Tonelli concerning convergence and of Tonelli and McShane concerning lower semicontinuity ([1], [2], [3], [4]). For example, if $\mathfrak{x}^n$ is a sequence of surfaces of class $\mathfrak{T}$ with uniformly bounded Dirichlet integral, $D[\mathfrak{x}^n] \leqq M$, and if $\mathfrak{x}^n$ converges uniformly to $\mathfrak{x}$, then $\mathfrak{x}$ is of class $\mathfrak{T}$. And $K[\mathfrak{x}] \leqq \liminf K[\mathfrak{x}^n]$ where K is any of the integrals A, D, F, I, J.

4. Convergence lemmas concerning the Dirichlet integral

LEMMA 4.1[4]. *Let $\mathfrak{x}^n$ be a sequence of surfaces converging uniformly to the surface $\mathfrak{x}$. Then*

$$D[\mathfrak{x}^n] \to D[\mathfrak{x}] \quad \textit{implies} \quad D[\mathfrak{x}^n - \mathfrak{x}] \to 0.$$

PROOF. Because

$$D[\mathfrak{x}^n - \mathfrak{x}] = D[\mathfrak{x}^n] - D[\mathfrak{x}] - 2D[\mathfrak{x}^n - \mathfrak{x}, \mathfrak{x}], \tag{4.1}$$

it is necessary to discuss

$$D[\mathfrak{x}^n - \mathfrak{x}, \mathfrak{x}] = \tfrac{1}{2} \iint (\mathfrak{x}_u^n - \mathfrak{x}_u)\mathfrak{x}_u + (\mathfrak{x}_v^n - \mathfrak{x}_v)\mathfrak{x}_v]\, du\, dv.$$

Consider the first part of the integral on the right hand side. Let $\epsilon > 0$ be arbitrary. Approximate $\mathfrak{x}_u$ in the mean by a polynomial surface $\mathfrak{p}$ such that

$$\iint (\mathfrak{x}_u - \mathfrak{p})^2\, du\, dv < \epsilon^2/16M$$

where M is an upper bound for $D[\mathfrak{x}^n]$, $n = 1, 2, \cdots$, and $D[\mathfrak{x}]$. The integral under consideration can be written

$$\tfrac{1}{2} \iint (\mathfrak{x}_u^n - \mathfrak{x}_u)(\mathfrak{x}_u - \mathfrak{p})\, du\, dv + \tfrac{1}{2} \iint (\mathfrak{x}_u^n - \mathfrak{x}_u)\mathfrak{p}\, du\, dv.$$

The first of these integrals is by Schwarz's inequality in absolute value $\leqq \left\{ D[\mathfrak{x}^n - \mathfrak{x}] \cdot \frac{\epsilon^2}{16M} \right\}^{1/2} \leqq \frac{\epsilon}{2}$. The second of these integrals is equal to, by integration by parts,

$$\int \mathfrak{p}(\mathfrak{x}^n - \mathfrak{x})\, dv - \iint \mathfrak{p}_u(\mathfrak{x}^n - \mathfrak{x})\, du\, dv$$

[4] Lemma 4.1 and its proof are easily generalized to include weak convergence of $\mathfrak{x}^n$ to $\mathfrak{x}$. Incidentally, (4.1) and (4.2) establish the lower semicontinuity of $D[\mathfrak{x}]$.

which is less than $\epsilon/2$ for all sufficiently large n. Hence, for these n's,

$$\left|\iint (\mathfrak{x}_u^n - \mathfrak{x}_u)\mathfrak{x}_u \, du \, dv\right| < \epsilon.$$

Similarly for $\iint (\mathfrak{x}_v^n - \mathfrak{x}_v)\mathfrak{x}_v \, du \, dv$. This proves

$$D[\mathfrak{x}^n - \mathfrak{x}, \mathfrak{x}] \to 0, \tag{4.2}$$

and with it the lemma.

LEMMA 4.2. *For any two surfaces* $\mathfrak{x}$ *and* $\mathfrak{x}'$,

$$|F[\mathfrak{x}'] - F[\mathfrak{x}]| \leqq 2k'(\sqrt{D[\mathfrak{x}']} + \sqrt{D[\mathfrak{x}]})\cdot\sqrt{D[\mathfrak{x}' - \mathfrak{x}]},$$

and

$$|I[\mathfrak{x}'] - I[\mathfrak{x}]| \leqq (2k' + k)(\sqrt{D[\mathfrak{x}']} + \sqrt{D[\mathfrak{x}]})\sqrt{D[\mathfrak{x}' - \mathfrak{x}]}.$$

PROOF. We have

$$\mathfrak{X}' - \mathfrak{X} = \mathfrak{x}_u' \times \mathfrak{x}_v' - \mathfrak{x}_u \times \mathfrak{x}_v = \mathfrak{x}_u' \times (\mathfrak{x}_v' - \mathfrak{x}_v) + (\mathfrak{x}_u' - \mathfrak{x}_u) \times \mathfrak{x}_v$$

and

$$\left(\iint \mathfrak{x}_u' \times (\mathfrak{x}_v' - \mathfrak{x}_v)\, du\, dv\right)^2 \leqq \iint \mathfrak{x}_u'^2 \, du \, dv \cdot \iint (\mathfrak{x}_v' - \mathfrak{x}_v)^2 \, du \, dv$$

$$\leqq 4D[\mathfrak{x}']\cdot D[\mathfrak{x}' - \mathfrak{x}].$$

Similarly for the other term in $\mathfrak{X}' - \mathfrak{X}$. Thus,

$$\iint |\mathfrak{X}' - \mathfrak{X}| \, du \, dv \leqq 2(\sqrt{D[\mathfrak{x}']} + \sqrt{D[\mathfrak{x}]})\cdot\sqrt{D[\mathfrak{x}' - \mathfrak{x}]}. \tag{4.3}$$

The inequality of the lemma is a consequence of (4.3) and property (2.c).

The inequality of the lemma involving the integral I requires in addition the triangle inequality

$$|\sqrt{D[\mathfrak{x}']} - \sqrt{D[\mathfrak{x}]}| \leqq \sqrt{D[\mathfrak{x}' - \mathfrak{x}]}. \tag{4.4}$$

LEMMA 4.3. *Let* $\mathfrak{x}^n$, $n = 1, 2, \cdots$, *and* $\mathfrak{x}$ *be surfaces such that*

$$D[\mathfrak{x}^n - \mathfrak{x}] \to 0 \text{ as } n \to \infty.$$

Then

$$K[\mathfrak{x}^n] \to K[\mathfrak{x}],$$

where K *is any one of the integrals* A, D, F, I, J.

PROOF. An immediate consequence of (4.4), Lemma 4.2, and (4.3).

II. Convex Sets of Surfaces

5. The uniqueness property for convex sets

The whole space of surfaces of class $\mathfrak{T}$ is unmanageable (for example, it is not compact). It will be decomposed into various mutually exclusive convex sets—in each set there will be a unique minimizing surface. This will permit a reduction of the whole space to the space of these minimizing surfaces (the *core* space).

Definition 5.1. A set $\mathfrak{K}$ of surfaces is said to be *convex* if

$$(1 - t)\mathfrak{x} + t\mathfrak{y} \quad \text{for} \quad 0 \leqq t \leqq 1$$

belongs to $\mathfrak{K}$ whenever $\mathfrak{x}$ and $\mathfrak{y}$ belong to $\mathfrak{K}$.

The significant theorem concerning convex sets is

Theorem 5.1. *Let $\mathfrak{K}$ be a convex set of surfaces such that there exists a surface $\mathfrak{x}$ minimizing $I[\mathfrak{x}]$ in the set $\mathfrak{K}$. Then $\mathfrak{x}$ is unique, except for possible translated surfaces $\mathfrak{x}$ + constant in $\mathfrak{K}$.*

Proof. Let $\mathfrak{x}'$ be any other surface in $\mathfrak{K}$ and consider the linear family of surfaces $\mathfrak{x}_t$, $0 \leqq t \leqq 1$, defined by

$$\mathfrak{x}_t = (1 - t)\mathfrak{x} + t\mathfrak{x}' = \mathfrak{x} + t\delta\mathfrak{x} \quad \text{where} \quad \delta\mathfrak{x} = \mathfrak{x}' - \mathfrak{x}.$$

Then

$$\mathfrak{X}_t = \frac{\partial \mathfrak{x}_t}{\partial u} \times \frac{\partial \mathfrak{x}_t}{\partial v} = \mathfrak{X} + t\delta\mathfrak{X} + t^2\delta^2\mathfrak{X}, \tag{5.1}$$

where

$$\delta\mathfrak{X} = \mathfrak{x}_u \times \delta\mathfrak{x}_v + \delta\mathfrak{x}_u \times \mathfrak{x}_v \quad \text{and} \quad \delta^2\mathfrak{X} = \delta\mathfrak{x}_u \times \delta\mathfrak{x}_v. \tag{5.2}$$

Also,

$$\frac{d\mathfrak{x}_t}{dt} = \delta\mathfrak{x} \quad \text{and} \quad \frac{d\mathfrak{X}_t}{dt} = \delta\mathfrak{X} + 2t\delta^2\mathfrak{X}. \tag{5.3}$$

We have $I[\mathfrak{x}_t] = \iint f(\mathfrak{X}_t)\, du\, dv + kD[\mathfrak{x}_t]$ and

$$\frac{dI[\mathfrak{x}_t]}{dt} = \iint f_{\mathfrak{X}_t} \cdot \frac{d\mathfrak{X}_t}{dt}\, du\, dv + 2kD\left[\mathfrak{x}_t, \frac{d\mathfrak{x}_t}{dt}\right], \tag{5.4}$$

where $D[\mathfrak{x}, \mathfrak{y}]$ is the 'cross' Dirichlet integral $\frac{1}{2}\iint (\mathfrak{x}_u\mathfrak{y}_u + \mathfrak{x}_v\mathfrak{y}_v)\, du\, dv$. The only difficulty with formula (5.4) is that $f_{\mathfrak{X}_t}$ is not defined at a point where $\mathfrak{X}_t = 0$. Because of this, we must distinguish in (5.4) between a right hand and a left hand derivative. For the right hand derivative $(d_+I[\mathfrak{x}_t])/dt$, define $f_{\mathfrak{X}_t}$ at a point where $\mathfrak{X}_t = 0$ to be

$$f_{\mathfrak{X}_t} = f_{\mathfrak{X}}\left(\frac{d\mathfrak{X}_t}{dt}\right). \tag{5.5a}$$

For a left hand derivative $(d_{-}I[\mathfrak{x}_t])/dt$, $f_{\mathfrak{X}_t}$ is defined at a point $\mathfrak{X}_t = 0$ to be

$$f_{\mathfrak{X}_t} = f_{\mathfrak{X}}\left(-\frac{d\mathfrak{X}_t}{dt}\right). \tag{5.5b}$$

In both cases, if $(d\mathfrak{X}_t)/dt$ is also zero at this point, the definition of $f_{\mathfrak{X}_t}$ in (5.4) is of no concern and it may be set equal to 0.

To prove (5.4) with the modifications (5.5a) or (5.5b), first note from (5.1) that

$$\mathfrak{X}_{t+\tau} = \mathfrak{X}_t + \tau\frac{d\mathfrak{X}_t}{dt} + \tau^2\delta^2\mathfrak{X},$$

and form the difference quotient

$$q = \frac{f(\mathfrak{X}_{t+\tau}) - f(\mathfrak{X}_t)}{\tau}.$$

By the theorem of the mean, in case τ is sufficiently small,

$$q = \left(\frac{d\mathfrak{X}_t}{dt} + \tau\delta^2\mathfrak{X}\right)f_{\mathfrak{X}} \tag{5.6}$$

where the derivative $f_{\mathfrak{X}}$ has the argument

$$\mathfrak{X}_t + \theta\tau\left(\frac{d\mathfrak{X}_t}{dt} + \tau\delta^2\mathfrak{X}\right) \quad \text{where} \quad 0 < \theta < 1. \tag{5.7}$$

In case $\mathfrak{X}_t \neq 0$, the limit of q as $\tau \to 0$ is

$$f_{\mathfrak{X}_t}\cdot\frac{d\mathfrak{X}_t}{dt}. \tag{5.8}$$

In case $\mathfrak{X}_t = 0$, the value of $f_{\mathfrak{X}}$ for the argument (5.7) is by positive homogoneity equal to its value for the argument $(d\mathfrak{X}_t)/dt + \tau\delta^2\mathfrak{X}$ or $-(d\mathfrak{X}_t)/dt - \tau\delta^2\mathfrak{X}$ according as $\tau > 0$ or $\tau < 0$. The limit of q as $\tau \to 0$ is therefore still (5.8), with the meaning (5.5a) or (5.5b).

Also, $|q| \leqq k' \,|\, (d\mathfrak{X}_t/dt \,|\, + k' \,|\, \delta^2\mathfrak{X}\,|$ which is integrable.[9] Since

$$\frac{F[\mathfrak{X}_{t+\tau}] - F[\mathfrak{X}_t]}{\tau} = \int\int q\, du\, dv,$$

it follows that $(dF[\mathfrak{x}_t])/dt$ is equal to the integral of the expression (5.8) with the agreement (5.5a) or (5.5b). The differentiation of the Dirichlet integral is clear, so that (5.4) with the understanding (5.5a) or (5.5b) is proved.

In particular,

$$\left.\frac{d_{+}I[\mathfrak{X}_t]}{dt}\right]_{t=0} = \int\int f_{\mathfrak{X}}\cdot\delta\mathfrak{X}\, du\, dv + 2kD[\mathfrak{x}, \delta\mathfrak{x}] \geqq 0 \tag{5.9}$$

since $I[\mathfrak{x}_t]$ has a minimum for $t = 0$ in the range $0 \leqq t \leqq 1$.

We have

$$F[\mathfrak{x}'] - F[\mathfrak{x}] \geqq \iint (\mathfrak{X}' - \mathfrak{X})\cdot f_{\mathfrak{X}}\, du\, dv$$

$$= \iint \delta\mathfrak{X}\cdot f_{\mathfrak{X}}\, du\, dv + \iint \delta^2\mathfrak{X}\cdot f_{\mathfrak{X}}\, du\, dv$$

$$\geqq \iint \delta\mathfrak{X}\cdot f_{\mathfrak{X}}\, du\, dv - k'D[\delta\mathfrak{x}]$$

using inequality 2(d). Also $D[\mathfrak{x}'] - D[\mathfrak{x}] = 2D[\mathfrak{x}, \delta\mathfrak{x}] + D[\delta\mathfrak{x}]$. Combining these two results,

$$I[\mathfrak{x}'] - I[\mathfrak{x}] \geqq \iint f_{\mathfrak{X}}\cdot\delta\mathfrak{X}\, du\, dv + 2kD[\mathfrak{x}, \delta\mathfrak{x}] + (k - k')D[\delta\mathfrak{x}] \geqq 0 \tag{5.10}$$

by (5.9).[5] Equality can hold only if $D[\delta\mathfrak{x}] = 0$, or $\delta\mathfrak{x} =$ constant. Theorem 5.1 is proved.

In the proof above we have merely used (5.9) as the property of the surface $\mathfrak{x}$. Therefore the following theorem has also been proved.

THEOREM 5.2. *The minimizing surface $\mathfrak{x}$ in the convex set $\mathfrak{K}$ is characterized by*

$$\iint f_{\mathfrak{X}}\cdot\delta\mathfrak{X}\, du\, dv + 2kD[\mathfrak{x}, \delta\mathfrak{x}] \geqq 0$$

for all $\delta\mathfrak{x}$ such that $\mathfrak{x} + \delta\mathfrak{x}$ is in $\mathfrak{K}$. (The vector $\delta\mathfrak{X}$ is defined in (5.2)).

Also, if $\mathfrak{x}$ and $\mathfrak{x}'$ are any two surfaces in $\mathfrak{K}$, relations (5.9) and (5.10) show that

$$I[\mathfrak{x}'] - I[\mathfrak{x}] \geqq \left.\frac{d_+ I[\mathfrak{x}_t]}{dt}\right]_{t=0}. \tag{5.11}$$

Selecting $\mathfrak{x}$ and $\mathfrak{x}'$ to be any two surfaces on the linear path $\mathfrak{x}_t$, $0 \leqq t \leqq 1$, (5.11) establishes the convexity of $I[\mathfrak{x}_t]$ as a function of t. Hence the following theorem.

THEOREM 5.3. *Let $\mathfrak{x}$ and $\mathfrak{x}'$ be any two surfaces in the convex set $\mathfrak{K}$, and form the linear path*

$$\mathfrak{x}_t = (1 - t)\mathfrak{x} + t\mathfrak{x}', \qquad 0 \leqq t \leqq 1$$

joining $\mathfrak{x}$ to $\mathfrak{x}'$. Then $I[\mathfrak{x}_t]$ is a convex function of t. In particular, if $\mathfrak{x}$ minimizes the integral I in the set $\mathfrak{K}$ then $I[\mathfrak{x}_t]$ is a monotonic increasing function of t.

Topologically, this theorem asserts that the set $\mathfrak{K}$ can be I-retracted into the single minimizing element $\mathfrak{x}$, i.e., retracted in such a way that the I-integral never increases along a path.

6. The set of surfaces with given boundary values

An example of a convex set is the set of all surfaces of class $\mathfrak{T}$ having prescribed boundary values. Designate such a set by the symbol $\mathfrak{B}$. We shall prove that

[5] Inequality (5.10) is more clearly understood if one notes that the hypotheses 2.2 and 2.4 imply the non-negativeness of the Weierstrass E-function for $I[\mathfrak{x}]$ when considered as a non-parametric problem. This is where hypothesis 2.4 is used.

a surface minimizing $I[\mathfrak{x}]$ in the set $\mathfrak{B}$ exists. For this purpose it is necessary to resort to a levelling process. Levelling will be performed in directions given by the stationary points of the function

$$g(\mathfrak{X}) = f(\mathfrak{X}) + f(-\mathfrak{X})$$

on the sphere $\mathfrak{X}^2 = 1$.[6]

It is easy to show that $g(\mathfrak{X})$ as a function on $\mathfrak{X}^2 = 1$ has three *linearly independent* stationary points.[7] Denote these stationary points by

$$(6.1) \qquad (\alpha_j, \beta_j, \gamma_j), \qquad j = 1, 2, 3.$$

(It is more convenient in this section to drop the vector notation.) At a stationary point of $g(X, Y, Z)$, the quantities g_X, g_Y, g_Z are proportional to X, Y, Z respectively. In particular if $(1, 0, 0)$ is a stationary point, $g_Y = g_Z = 0$ or

$$(6.2) \qquad \begin{cases} f_Y(1, 0, 0) = f_Y(-1, 0, 0) \\ f_Z(1, 0, 0) = f_Z(-1, 0, 0). \end{cases}$$

In place of the components x, y, z of a surface $\mathfrak{x}$, we shall use the functions

$$(6.3) \qquad \xi_j(u, v) = \alpha_j x(u, v) + \beta_j y(u, v) + \gamma_j z(u, v), \qquad j = 1, 2, 3.$$

These functions $\xi_j(u, v)$, $j = 1, 2, 3$, will be called the *affine components* of $\mathfrak{x}(u, v)$. Since the points (6.1) are linearly independent, the equations (6.3) can be solved for x, y, z in terms of ξ_1, ξ_2, ξ_3:

$$(6.4) \qquad x(u, v) = a_1\xi_1(u, v) + a_2\xi_2(u, v) + a_3\xi_3(u, v) \quad \text{etc.}$$

A surface is uniquely determined by stating its affine components. And specifying the boundary values of $\mathfrak{x}$ is equivalent to specifying the boundary values of its affine components.

Lemma 6.1. *There is a constant c such that*

$$\tfrac{1}{3}(D[\xi_1] + D[\xi_2] + D[\xi_3]) \leqq D[\mathfrak{x}] \leqq c(D[\xi_1] + D[\xi_2] + D[\xi_3]),$$

where ξ_1, ξ_2, ξ_3 are the affine components of the surface $\mathfrak{x}$.

Proof. The first inequality is obtained by applying the triangle inequality for the Dirichlet integral to (6.3):

$$\sqrt{D[\xi_j]} \leqq |\alpha_j| \sqrt{D[x]} + |\beta_j| \sqrt{D[y]} + |\gamma_j| \sqrt{D[z]} \leqq \sqrt{D[\mathfrak{x}]}$$

[6] Levelling is used, among other places, in [2] and [4]. For the parametric problem, the function $g(\mathfrak{X})$ is used in [4].

[7] Proof. $g(\mathfrak{X})$ satisfies $g(-\mathfrak{X}) = g(\mathfrak{X})$. Let P, p be distinct points on the unit sphere $\mathfrak{X}^2 = 1$ where g assumes its maximum and minimum values respectively. Minimize g on each great circle through P, and determine a great circle Σ for which this minimum is the largest possible. A point Q on Σ at which g attains its minimum on Σ is a stationary point of $g(\mathfrak{X})$ on $\mathfrak{X}^2 = 1$. Such a great circle Σ and such a point Q can be selected so that the points P, p, Q are linearly independent. Q.E.D.

since $\alpha_i^2 + \beta_i^2 + \gamma_i^2 = 1$. The second inequality of the lemma follows from the triangle inequality applied to (6.4):

$$\sqrt{D[x]} \leqq |a_1| \sqrt{D[\xi_1]} + |a_2| \sqrt{D[\xi_2]} + |a_3| \sqrt{D[\xi_3]}$$
$$\leqq (a_1^2 + a_2^2 + a_3^2)^{1/2} \cdot (D[\xi_1] + D[\xi_2] + D[\xi_3])^{1/2}$$

etc.

Returning to the problem of minimizing $I[\mathfrak{x}]$ in the set $\mathfrak{B}$, let i be the greatest lower bound of $I[\mathfrak{x}]$ for all $\mathfrak{x}$ in $\mathfrak{B}$. We shall level any surface $\mathfrak{x}$ in $\mathfrak{B}$ to yield a surface in $\mathfrak{B}$ the affine components of which have arbitrarily small monotonic deficiency.

LEMMA 6.2. *Suppose that* (1, 0, 0) *is a stationary point of* $g(X, Y, Z)$ *on* $X^2 + Y^2 + Z^2 = 1$. *Given any surface* $\mathfrak{x}$ *in* $\mathfrak{B}$ *and any positive* δ, *there is a surface* $\mathfrak{x}'$ *in* $\mathfrak{B}$ *with x-component* x' *of monotonic deficiency* $\leqq \delta$, *and satisfying*

$$D[x' - x] \leqq (I[\mathfrak{x}] - i)/k.$$

PROOF. Suppose that $\mathfrak{x}$ is not level relative to some of the planes $x = m\delta/2$, $m = 0, \pm 1, \pm 2, \cdots$, i.e., there exists a region G of the (u, v)-unit circle with boundary G^* such that $x(u, v) = m\delta/2$ on G^* for some integer m, while $x(u, v) \not\equiv m\delta/2$ inside G. Consider all such regions and retain only maximal G's, deleting all others. The retained regions G_i are non-overlapping. Define the surface $\mathfrak{x}'$ as follows:

$$\begin{cases} x' = m_i\delta/2 \text{ in each } G_i, \ x' = x \text{ outside all the } G_i\text{'s}; \\ y' = y, \ z' = z. \end{cases}$$

The surface $\mathfrak{x}'$ is in $\mathfrak{B}$ and is level relative to all the planes $x = m\delta/2$, $m = 0, \pm 1, \pm 2, \cdots$. The latter property means that the monotonic deficiency of x' is $\leqq \delta$.

In each region G_i

$$F_{G_i}[\mathfrak{x}] - F_{G_i}[\mathfrak{x}'] \geqq \iint_{G_i} \{Yf_Y(X, 0, 0) + Zf_Z(X, 0, 0)\}\, du\, dv.$$

But $f_Y(X, 0, 0) = f_Y(\pm 1, 0, 0) = c_1$, $f_Z(X, 0, 0) = f_Z(\pm 1, 0, 0) = c_2$, by positive homogeneity and (6.2). Therefore,

$$F_{G_i}[\mathfrak{x}] - F_{G_i}[\mathfrak{x}'] \geqq c_1 \iint_{G_i} Y\, du\, dv + c_2 \iint_{G_i} Z\, du\, dv = 0, \tag{6.5}$$

since $\iint_{G_i} Y\, du\, dv = \int_{G_i^*} z\, dx$, $\int_{G_i} Z\, du\, dv = -\int_{G_i^*} y\, dx$, and x = constant on G_i^*.[8] Since $\mathfrak{x} = \mathfrak{x}'$ outside the G_i, (6.5) yields

$$F[\mathfrak{x}] - F[\mathfrak{x}'] \geqq 0. \tag{6.6}$$

[8] See [4], p. 558 for this argument.

Concerning the Dirichlet integral, we have

$$D_{G_i}[\mathfrak{x}] - D_{G_i}[\mathfrak{x}'] = D_{G_i}[x - x'],$$

so that

$$D[\mathfrak{x}] - D[\mathfrak{x}'] = D[x - x']. \tag{6.7}$$

Combining (6.6) and (6.7),

$$I[\mathfrak{x}] - I[\mathfrak{x}'] \geqq kD[x - x'],$$

and the inequality of the lemma is proved by noting that $I[\mathfrak{x}'] \geqq i$.

LEMMA 6.3. *Let $\mathfrak{x}$ be any surface in $\mathfrak{B}$, and δ any positive number. There is a surface $\mathfrak{y}$ in $\mathfrak{B}$ with affine components of monotonic deficiency $\leqq \delta$, and satisfying*

$$D[\mathfrak{y} - \mathfrak{x}] \leqq 3c(I[\mathfrak{x}] - i)/k,$$

where c is the constant of Lemma 6.1.

PROOF. Consider first the stationary point $(\alpha_1, \beta_1, \gamma_1)$ of g. Perform the orthogonal transformation

$$\begin{cases} \xi_1 = \alpha_1 x + \beta_1 y + \gamma_1 z \\ y^* = \alpha_1' x + \beta_1' y + \gamma_1' z \\ z^* = \alpha_1'' x + \beta_1'' y + \gamma_1'' z \end{cases}$$

The surface $\mathfrak{x}$ becomes a surface $\mathfrak{x}^*$ with coordinate functions ξ_1, y^*, z^*; $X^* = \alpha_1 X + \beta_1 Y + \gamma_1 Z$ etc., $E^* = E$, etc., and $I[\mathfrak{x}] = \iint f^*(X^*, Y^*, Z^*)\, du\, dv + kD[\mathfrak{x}^*]$, where $f^*(X^*, Y^*, Z^*) = f(X, Y, Z)$. Also, $g^*(\mathfrak{X}^*) = g(\mathfrak{X})$, and a stationary point of g^* is $(X^*, Y^*, Z^*) = (1, 0, 0)$. Lemma 6.2 applied to $\mathfrak{x}^*$ establishes the existence of a function ξ_1' having the desired boundary values, with monotonic deficiency $\leqq \delta$, and $D[\xi_1' - \xi_1] \leqq (I[\mathfrak{x}] - i)/k$.

Likewise, using the other stationary points of g, there exist functions ξ_2', ξ_3' enjoying analogous properties.

Let $\mathfrak{y}$ be the surface with ξ_1', ξ_2', ξ_3' as its affine components. It is the surface required by the lemma, in view of lemma 6.1.[9]

It is now easy to prove the main theorem of this section.

THEOREM 6.1. *There is a unique surface $\mathfrak{x}$ minimizing $I[\mathfrak{x}]$ in the set $\mathfrak{B}$ of all surfaces having prescribed boundary values. Its affine components ξ_1, ξ_2, ξ_3 are monotonic functions.*

PROOF. Let $\mathfrak{x}^n$ be a minimizing sequence, $I[\mathfrak{x}^n] \to i$. For each $\mathfrak{x}^n$, construct the surface $\mathfrak{y}^n$ of lemma 6.3 with $\delta = 1/n$. Lemma 6.3 asserts that

$$D[\mathfrak{y}^n - \mathfrak{x}^n] \to 0,$$

---

[9] It would seem that the most direct way to prove Lemma 6.3 and Theorem 6.1 is to level with respect to ξ_1 then with respect to ξ_2 and ξ_3. But this method fails here because a levelling relative to ξ_2 may destroy the previously obtained levelling relative to ξ_1.

and Lemma 4.2 that

$$I[\mathfrak{y}^n] - I[\mathfrak{x}^n] \to 0.$$

Therefore $\mathfrak{y}^n$ is also a minimizing sequence.

By the properties of the affine components of $\mathfrak{y}^n$ stated in Lemma 6.3, a subsequence of the $\mathfrak{y}^n$'s has affine components which converge uniformly to monotonic functions. This subsequence of the $\mathfrak{y}^n$'s therefore converges uniformly to a surface $\mathfrak{x}$ in $\mathfrak{B}$ with these monotonic affine components. The lower semicontinuity of the I-functional shows that $\mathfrak{x}$ is a solution to the minimum problem. Uniqueness follows from theorem 5.1. Q.E.D.

Theorem 5.2 becomes

THEOREM 6.2. *The surface $\mathfrak{x}$ minimizing $I[\mathfrak{x}]$ in the set $\mathfrak{B}$ of all surfaces with prescribed boundary values is characterized by the variational condition*

$$(V_1) \qquad \iint f_{\mathfrak{X}} \cdot \delta\mathfrak{X}\, du\, dv + 2kD[\mathfrak{x}, \delta\mathfrak{x}] \geqq 0$$

for all surfaces $\delta\mathfrak{x}$ of class $\mathfrak{T}$ with boundary values 0. (If the points where $\mathfrak{X} = 0$ formed a set of measure 0, the inequality would become an equality.)

A surface $\mathfrak{x}$ satisfying (V_1) will be called an *I-surface.* It minimizes $I[\mathfrak{x}]$ among all surfaces having the same boundary values.

III. CONVERGENCE, ISOPERIMETRIC AND CONTINUITY PROPERTIES OF I-SURFACES

7. Elementary properties of I-surfaces

We repeat the definition: an I-surface is a surface of class $\mathfrak{T}$ which satisfies the variational condition (V_1) stated above in Theorem 6.2.

For the special case of the Plateau problem, when $f \equiv 0$, I-surfaces are potential surfaces. This part III is devoted to the proof of theorems concerning I-surfaces which are pertinent to our present use and are analogous to corresponding theorems for potential surfaces. The major device for proving these theorems is a comparison between I-surfaces and potential surfaces based on Lemma 8.1.

7.1. If $\mathfrak{x}$ is an I-surface, so is $\mathfrak{x} + \mathfrak{c}$.

7.2. The property of being an I-surface is invariant under direct conformal mapping of the parameter (u, v)-plane. This follows from the invariance of the integral I under direct conformal mapping.

7.3. The fact that the affine components ξ_1, ξ_2, ξ_3 are monotonic can be expressed most aptly in terms of the concept of the *monotonic quotient* of a surface $\mathfrak{x}(u, v)$. Consider any subregion G of the unit circle, with boundary G^*, and form the quotient $q_G = \dfrac{\text{diameter of } \mathfrak{x}(u, v) \text{ over } G}{\text{diameter of } \mathfrak{x}(u, v) \text{ over } G^*}$. (If both numerator and denominator are 0, q_G is taken as 1.) The *monotonic quotient* Q of the surface $\mathfrak{x}(u, v)$ is the least upper bound of q_G for all subregions G. If Q is $\neq \infty$, the surface $\mathfrak{x}(u, v)$ is said to have a bounded monotonic quotient. (If $\mathfrak{x}(u, v)$ is a saddle surface, $Q = 1$.)

Now, the totality of I-surfaces has uniformly bounded monotonic quotient. For, from the equation (6.3) there follows the fact that the oscillation of each ξ_j on G^* is at most equal to the diameter of $\mathfrak{x}$ on G^*. By the monotonicity of ξ_j, therefore, the oscillation of each ξ_j in G is at most equal to the diameter of $\mathfrak{x}$ on G^*. This and equation (6.4) show that the diameter of $\mathfrak{x}$ in G is at most equal to $3c$ times the diameter of $\mathfrak{x}$ on G^*, where c is the constant derived in Lemma 6.1. Thus, the monotonic quotient of $\mathfrak{x}$ is at most $3c$.

7.4. A significant use of 7.3 is contained in the following easily proved result. A set of surfaces of class $\mathfrak{T}$ with uniformly bounded Dirichlet integral and with uniformly bounded monotonic quotient is equicontinuous in every closed domain interior to the unit circle.

8. Comparison with potential surfaces

LEMMA 8.1. *Let $\mathfrak{x}(u, v)$ be an I-surface, and G any region of the (u, v)-unit circle with boundary G^*. Let $\mathfrak{p}(u, v)$ be the potential function inside G having the same boundary values as $\mathfrak{x}(u, v)$ on G^*. Then*

$$F_G[\mathfrak{x}] \leqq F_G[\mathfrak{p}] - kD_G[\mathfrak{x} - \mathfrak{p}]. \tag{8.1}$$

In particular,

$$F_G[\mathfrak{x}] \leqq F_G[\mathfrak{p}] \leqq k'A_G[\mathfrak{p}], \tag{8.2}$$

and

$$D_G[\mathfrak{x} - \mathfrak{p}] \leqq \frac{k'}{k} A_G[\mathfrak{p}]. \tag{8.3}$$

PROOF. Define the surface $\mathfrak{x}'(u, v)$ by setting $\mathfrak{x}' = \mathfrak{p}$ inside G and $\mathfrak{x}' = \mathfrak{x}$ outside G. It is clear that $\mathfrak{x}'$ is of class $\mathfrak{T}$. Since $\mathfrak{x}'$ has the same boundary values as $\mathfrak{x}$, and $\mathfrak{x}$ is an I-surface,

$$I[\mathfrak{x}] \leqq I[\mathfrak{x}']$$

or

$$F_G[\mathfrak{x}] + kD_G[\mathfrak{x}] \leqq F_G[\mathfrak{p}] + kD_G[\mathfrak{p}]. \tag{8.4}$$

But

$$D_G[\mathfrak{p}] = D_G[\mathfrak{x}] - D_G[\mathfrak{x} - \mathfrak{p}],$$

and substitution in (8.4) yields (8.1). q.e.d.

We shall find frequent occasion to use the potential surface defined over the unit circle having the same boundary values as $\mathfrak{x}$. This will be called the *potential surface associated* with $\mathfrak{x}$.

9. An isoperimetric inequality

By an isoperimetric inequality will be meant an inequality of the type

$$A[\mathfrak{x}] \leqq KL[\mathfrak{x}]^2$$

applicable to a specified class of surfaces, where K is a constant for the class and $L[\mathfrak{x}]$ is the length of the boundary of $\mathfrak{x}$. We shall establish such an inequality for I-surfaces, using one for potential surfaces.

Beckenbach and Radó have established the usual isoperimetric inequality $A \leqq L^2/4\pi$ for sufficiently regular surfaces of non-positive curvature at each point.[10] Since potential surfaces have non-positive curvature, all that remains to obtain $A \leqq L^2/4\pi$ is to establish the possibility of a conformal representation of a potential surface which does not lead to singular points. There are strong indications that this can be done.

At any rate, a recent work of Morse and Tompkins implicitly contains an isoperimetric inequality for potential surfaces ([20]). They do not state any isoperimetric inequality, but their procedure yields $A \leqq KL^2$ with some constant K. By a slight refinement in their methods, it is possible to establish $A \leqq L^2/4$. In this paper, we shall use

$$A \leqq KL^2 \quad \text{for potential surfaces,} \tag{9.1}$$

where we know that $1/4\pi \leqq K \leqq 1/4$ and we expect the best K to be $1/4\pi$.[10a]

THEOREM 9.1. *For an I-surface $\mathfrak{x}$ bounded by a curve of length L, we have*

$$A[\mathfrak{x}] \leqq \frac{k'K}{m} L^2$$

and

$$J[\mathfrak{x}] \leqq \left(1 + \frac{k}{m}\right) k'KL^2.$$

PROOF. Let $\mathfrak{p}$ be the potential surface associated with $\mathfrak{x}$. Inequality (8.2) in Lemma 8.1 yields

$$F[\mathfrak{x}] \leqq k'A[\mathfrak{p}] \leqq k'KL^2.$$

The inequality $mA[\mathfrak{x}] \leqq F[\mathfrak{x}]$ (see 2.3) gives the first inequality of the theorem. The second inequality is obtained from $J[\mathfrak{x}] = F[\mathfrak{x}] + kA[\mathfrak{x}]$.

10. Convergence theorems

THEOREM 10.1. *Let $\{\mathfrak{x}\}$ be any set of I-surfaces with boundary curves of length $\leqq L$. Then the set $\{D_S[\mathfrak{x}]\}$ for any closed interior domain S is uniformly bounded. In particular, the set $\{\mathfrak{x}\}$ is equicontinuous in any closed interior domain S.*

PROOF. Let $\{\mathfrak{p}\}$ be the set of potential surfaces associated with $\{\mathfrak{x}\}$. It is well known that the set $\{D_S[\mathfrak{p}]\}$ is uniformly bounded; in fact, $D_S[\mathfrak{p}] \leqq L^2/\pi d^2$

---

[10] E. F. Beckenbach and T. Radó, Subharmonic functions and surfaces of negative curvature, Trans. Amer. Math. Soc., 35, 662–674 (1933).

[10a] In the meantime, the author has obtained a completely elementary proof of $A \leqq L^2/4\pi$. Thus K can be set equal to $1/4\pi$.

where d is the minimum distance of points of S from the circumference of the unit circle. By inequality (8.3) in lemma 8.1,

$$D[\mathfrak{x} - \mathfrak{y}] \leqq \frac{k'}{k} A[\mathfrak{y}] \leqq \frac{k'K}{k} L^2.$$

The triangle inequality shows that

$$D_S[\mathfrak{x}] \leqq (\sqrt{D_S[\mathfrak{y}]} + \sqrt{D_S[\mathfrak{x} - \mathfrak{y}]})^2 \leqq \left(\frac{1}{\sqrt{\pi}\, d} + \sqrt{\frac{k'K}{m}}\right)^2 L^2,$$

so that the set $\{D_S[\mathfrak{x}]\}$ is uniformly bounded.

Equicontinuity follows from 7.3 and 7.4.

The remaining theorems discuss the limit surface of a sequence of I-surfaces.

Theorem 10.2. *Let $\mathfrak{x}^n$ be a sequence of I-surfaces converging to a surface $\mathfrak{x}$ uniformly in every closed interior domain S. Suppose also that $D_S[\mathfrak{x}^n] \leqq M(S)$ for all n and every S. Then*

$$I_S[\mathfrak{x}^n] \to I_S[\mathfrak{x}].$$

Also,

$$D_S[\mathfrak{x}^n - \mathfrak{x}] \to 0.$$

Proof. Since $\mathfrak{x}^n$ is an I-surface, condition (V_1) of theorem 6.2 is satisfied for $\mathfrak{x}^n$. Select $\rho < 1$, and a ρ^n soon to be determined between ρ and $\bar{\rho} = (1 + \rho)/2$. Select $\delta\mathfrak{x}^n$ as follows:

$$\delta\mathfrak{x}^n = \begin{cases} \mathfrak{x} - \mathfrak{x}^n & \text{in } C_\rho \\ \dfrac{\rho^n - r}{\rho^n - \rho} \cdot (\mathfrak{x} - \mathfrak{x}^n) & \text{in } C_{\rho\rho^n} \\ 0 & \text{in } C_{\rho^n 1} \end{cases} \tag{10.1}$$

In C_ρ we have by (5.1) and (5.2),

$$\delta\mathfrak{X}^n = \mathfrak{X} - \mathfrak{X}^n - \delta^2\mathfrak{X}^n$$

where $\mathfrak{X}$, $\mathfrak{X}^n$ are the normal vectors for the surfaces $\mathfrak{x}$, $\mathfrak{x}^n$. The variational condition (V_1) for the surface $\mathfrak{x}^n$, together with $2D[\mathfrak{x}^n, \mathfrak{x} - \mathfrak{x}^n] = D[\mathfrak{x}] - D[\mathfrak{x}^n] - D[\mathfrak{x} - \mathfrak{x}^n]$, yields

$$\iint_{C_\rho} f_{\mathfrak{X}^n} \cdot (\mathfrak{X} - \mathfrak{X}^n - \delta^2\mathfrak{X}^n)\, du\, dv + k(D_{C_\rho}[\mathfrak{x}] - D_{C_\rho}[\mathfrak{x}^n] - D_{C_\rho}[\mathfrak{x} - \mathfrak{x}^n]) \tag{10.2}$$

$$\geqq - \iint_{C_{\rho\rho^n}} f_{\mathfrak{X}^n} \cdot \delta\mathfrak{X}^n\, du\, dv - 2kD_{C_{\rho\rho^n}}[\mathfrak{x}^n, \delta\mathfrak{x}^n].$$

We shall use (10.2) to estimate $I_{C_\rho}[\mathfrak{x}] - I_{C_\rho}[\mathfrak{x}^n]$. We have

$$I_{C_\rho}[\mathfrak{x}] - I_{C_\rho}[\mathfrak{x}^n] \geqq \iint_{C_\rho} (\mathfrak{X} - \mathfrak{X}^n)\cdot f_{\mathfrak{X}^n}\, du\, dv + k(D_{C_\rho}[\mathfrak{x}] - D_{C_\rho}[\mathfrak{x}^n])$$

$$\geqq \iint_{C_\rho} f_{\mathfrak{X}^n}\cdot \delta^2 \mathfrak{X}^n\, du\, dv + kD_{C_\rho}[\mathfrak{x} - \mathfrak{x}^n]$$

$$- \iint_{C_{\rho\rho^n}} f_{\mathfrak{X}^n}\cdot \delta\mathfrak{X}^n\, du\, dv - 2kD_{C_{\rho\rho^n}}[\mathfrak{x}^n, \delta\mathfrak{x}^n],$$

by (10.2). The sum of the first two terms of the right hand member is $\geqq 0$, by (2.d) and (2.4). Therefore,

$$I_{C_\rho}[\mathfrak{x}] - I_{C_\rho}[\mathfrak{x}^n] \geqq -\iint_{C_{\rho\rho^n}} f_{\mathfrak{X}^n}\cdot \delta\mathfrak{X}^n\, du\, dv - 2kD_{C_{\rho\rho^n}}[\mathfrak{x}^n, \delta\mathfrak{x}^n]. \tag{10.3}$$

In $C_{\rho\rho^n}$, by (10.1),

$$(\delta\mathfrak{x}^n)_u = \frac{\rho^n - r}{\rho^n - \rho}(\mathfrak{x}_u - \mathfrak{x}^n_u) - \frac{\cos\theta}{\rho^n - \rho}(\mathfrak{x} - \mathfrak{x}^n)$$

or

$$|(\delta\mathfrak{x}^n)_u| \leqq |\mathfrak{x}_u - \mathfrak{x}^n_u| + \frac{|\mathfrak{x} - \mathfrak{x}^n|}{\rho^n - \rho},$$

with similar results for $(\delta\mathfrak{x}^n)_v$. Consequently,

$$\begin{aligned} D_{C_{\rho\rho^n}}[\delta\mathfrak{x}^n] &\leqq D_{C_{\rho\rho^n}}[\mathfrak{x} - \mathfrak{x}^n] + \\ &+ 2\,\frac{\max.|\mathfrak{x} - \mathfrak{x}^n|}{\rho^n - \rho}\,(2\pi(\rho^n - \rho)D_{C_{\rho\rho^n}}[\mathfrak{x} - \mathfrak{x}^n])^{\frac{1}{2}} \\ &+ \left(\frac{\max.|\mathfrak{x} - \mathfrak{x}^n|}{\rho^n - \rho}\right)^2\cdot 2\pi(\rho^n - \rho), \end{aligned} \tag{10.4}$$

where the maximum is taken in the domain $C_{\bar\rho}$. Now select ρ^n as follows:

$$\rho^n - \rho = \max|\mathfrak{x} - \mathfrak{x}^n|,$$

which approaches 0 as $n \to \infty$. By hypothesis, $D_{C_{\rho\rho^n}}[\mathfrak{x} - \mathfrak{x}^n]$ is uniformly bounded, and (10.4) results in

$$\limsup D_{C_{\rho\rho^n}}[\delta\mathfrak{x}^n] \leqq \limsup D_{C_{\rho\rho^n}}[\mathfrak{x} - \mathfrak{x}^n]. \tag{10.5}$$

It is a simple theorem, using merely the non-negative character of the integrand of the Dirichlet integral, that for any S there is a $\rho < 1$ and a subsequence of the n's such that

$$S \subset C_\rho \quad\text{and}\quad D_{C_{\rho\rho^n}}[\mathfrak{x} - \mathfrak{x}^n] \to 0$$

for this subsequence (and consider this subsequence henceforth).[11] As a result of inequality (10.5),

$$D_{C_{\rho\rho^n}}[\delta\mathfrak{x}^n] \to 0. \tag{10.6}$$

And (10.6) together with the uniform boundedness of $D_{C_{\rho\rho^n}}[\mathfrak{x}^n]$ show that the terms appearing on the right hand side of (10.3) approach 0. Consequently,

$$I_{C_\rho}[\mathfrak{x}] \geqq \limsup I_{C_\rho}[\mathfrak{x}^n].$$

But by lower semicontinuity,

$$I_{C_\rho}[\mathfrak{x}] \leqq \liminf I_{C_\rho}[\mathfrak{x}^n].$$

Finally,

$$I_{C_\rho}[\mathfrak{x}] = \lim I_{C_\rho}[\mathfrak{x}^n].$$

And this implies

$$I_S[\mathfrak{x}] = \lim I_S[\mathfrak{x}^n]. \tag{10.7}$$

To obtain the second part of the theorem, the combination of $I[\mathfrak{x}] = F[\mathfrak{x}] + kD[\mathfrak{x}]$, the lower semicontinuity of $F[\mathfrak{x}]$ and of $D[\mathfrak{x}]$, and (10.7) give

$$D_S[\mathfrak{x}] = \lim D_S[\mathfrak{x}^n].$$

And Lemma 4.1 yields $D_S[\mathfrak{x} - \mathfrak{x}^n] \to 0$.

THEOREM 10.3 *Under the same hypotheses as Theorem* 10.2, $\mathfrak{x}$ *is an I-surface in every closed interior domain S.*

PROOF. Let $\delta\mathfrak{x}$ be any surface of class $\mathfrak{T}$ which is identically 0 in the strip outside S. We shall show that the left hand side of (V_1) in Theorem 6.2 is upper semi continuous.[12]

As a result of the second part of theorem 10.2, we have

$$\iint |\mathfrak{X}^n - \mathfrak{X}| \, du \, dv \to 0, \tag{10.8}$$

and

$$\iint |\delta\mathfrak{X}^n - \delta\mathfrak{X}| \, du \, dv \to 0 \tag{10.9}$$

---

[11] Proof. Let $m(\rho) = \text{lim. inf. } D_{C_{\rho\rho^n}}[\mathfrak{x}^n - \mathfrak{x}]$. Let $\rho_1, \rho_2, \cdots, \rho_q$ be distinct values of ρ for which $m(\rho) \geqq a > 0$. For sufficiently large n the annular rings $C_{\rho_i\rho_i^n}$, $i = 1, 2, \cdots, q$, are non-overlapping, and $qa \leqq M$, where M is a uniform bound for $D[\mathfrak{x}^n - \mathfrak{x}]$. Thus $q \leqq M/a$ and there are only a finite number of ρ's for which $m(\rho) \geqq a$, hence at most an enumerable number of ρ's for which $m(\rho) \neq 0$. The theorem is now immediate.

[12] If the points where $\mathfrak{X} = 0$ formed a set of measure 0, the left hand side of (V_1) would be continuous. It is the possibility that this set might not be of measure 0 that complicates the proof.

as in the proof of (4.3). Let σ be the point set of S where $\mathfrak{X} = 0$; σ^n the point set of S outside σ where $\mathfrak{X}^n = 0$. Concerning σ^n, equation (10.8) gives

$$\iint_{\sigma^n} |\mathfrak{X}| \, du \, dv \to 0,$$

or

$$\text{meas. } \sigma^n \to 0. \tag{10.10}$$

Now, over $S - \sigma$,

$$\iint_{S-\sigma} f_{\mathfrak{x}^n} \cdot \delta\mathfrak{X}^n \, du \, dv \to \iint_{S-\sigma} f_{\mathfrak{x}} \cdot \delta\mathfrak{X} \, du \, dv. \tag{10.11}$$

This follows by writing

$$\iint_{S-\sigma} \{f_{\mathfrak{x}^n} \cdot \delta\mathfrak{X}^n - f_{\mathfrak{x}} \cdot \delta\mathfrak{X}\} \, du \, dv = \iint_{S-\sigma} f_{\mathfrak{x}^n} \cdot (\delta\mathfrak{X}^n - \delta\mathfrak{X}) \, du \, dv$$

$$+ \iint_{S-\sigma-\sigma n} (f_{\mathfrak{x}^n} - f_{\mathfrak{x}}) \cdot \delta\mathfrak{X} \, du \, dv + \iint_{\sigma^n} (f_{\mathfrak{x}^n} - f_{\mathfrak{x}}) \delta\mathfrak{X} \, du \, dv,$$

and noting that each of these terms approach 0 by (10.9), (10.8) and (10.10) respectively, and the Lebesgue theorem on the passage to the limit under the sign of integration.

Also, over the set σ,

$$\left\{\begin{aligned} \iint_{\sigma} f_{\mathfrak{x}^n} \cdot \delta\mathfrak{X}^n \, du \, dv &\leqq \iint_{\sigma} f(\delta\mathfrak{X}^n) \, du \, dv \to \iint_{\sigma} f(\delta\mathfrak{X}) \, du \, dv \\ &= \iint_{\sigma} f_{\mathfrak{x}} \cdot \delta\mathfrak{X} \, du \, dv \end{aligned}\right. \tag{10.12}$$

since the argument for $f_{\mathfrak{x}}$ in the set σ is $\delta\mathfrak{X}$.

Relations (10.11) and (10.12) lead to the desired result:

$$\iint f_{\mathfrak{x}} \cdot \delta\mathfrak{X} \, du \, dv \geqq \limsup \iint f_{\mathfrak{x}^n} \cdot \delta\mathfrak{X}^n \, du \, dv.$$

Of course, $D_S[\mathfrak{x}, \delta\mathfrak{x}] = \lim D_S[\mathfrak{x}^n, \delta\mathfrak{x}]$. Thus we have established the upper semicontinuity of the left hand side of (V_1) in Theorem 6.2.

Because each $\mathfrak{x}^n$ is an I-surface, there follows

$$\iint f_{\mathfrak{x}} \cdot \delta\mathfrak{X} \, du \, dv + 2kD[\mathfrak{x}, \delta\mathfrak{x}] \geqq 0$$

for every surface $\delta\mathfrak{x}$ with values 0 on the boundary of S. This is (V_1), so that $\mathfrak{x}$ is an I-surface in S. Q.E.D.

In the above theorem, the reason that we do not assert that $\mathfrak{x}$ is an I-surface over the whole unit circle is that the surface $\mathfrak{x}$ may fail to be of class $\mathfrak{X}$ on two counts: $D[\mathfrak{x}]$ when taken over the whole unit circle may not exist, and $\mathfrak{x}$ may not be continuous over the closed unit circle.

The next theorem discusses the boundary values of the limit surface of a sequence of I-surfaces.

THEOREM 10.4. *Let $\mathfrak{x}^n$ be a sequence of I-surfaces, bounded by the curves Γ^n, converging uniformly in every closed interior domain to a surface $\mathfrak{x}$ whose Dirichlet integral exists. Suppose that the curves Γ^n have uniformly bounded lengths and converge (in the Frechet sense) to a curve Γ. Then the surface $\mathfrak{x}$ is continuous in the closed unit circle and has boundary values lying monotonically on the curve Γ.*

PROOF. Select a proper representation $\mathfrak{g}^n(\varphi)$, $0 \leqq \varphi \leqq 2\pi$, of each Γ^n in such a way that $\mathfrak{g}^n(\varphi)$ converges uniformly to a proper representation $\mathfrak{g}(\varphi)$ of Γ. This is possible by virtue of the Frechet convergence of Γ^n to Γ. The boundary values $\mathfrak{x}^n(\theta)$ of $\mathfrak{x}^n$ have the form

$$\mathfrak{g}^n(\lambda^n(\theta))$$

where $\lambda^n(\theta)$ is a continuous monotonic function of θ, and $\lambda^n(\theta + 2\pi) = \lambda^n(\theta) + 2\pi$.

It is a classic theorem (see [10]) that a sequence of bounded monotonic functions contains a subsequence which converges. Applying this to $\lambda^n(\theta)$, and denoting the resulting subsequence again by $\lambda^n(\theta)$, we have $\lambda^n(\theta)$ converging to a monotonic function $\lambda(\theta)$ (which may have jump discontinuities). This subsequence has an interesting property which is less restrictive than equicontinuity and will be called weak-equicontinuity.

A sequence of functions $f^n(P)$ defined over any metric space will be called *weakly-equicontinuous* at a point P_0 if for any positive ϵ there is a positive δ with the following property: for any *closed connected* set Δ in the δ-neighborhood of P_0 *which does not include* P_0, an N (depending on Δ) can be found such that the oscillation of $f^n(P)$ on Δ is less than ϵ for all $n \geqq N$.

The modifier "weak" refers to the exclusion of the point P_0 from the set Δ. It is easy to prove that the sequence $\lambda^n(\theta)$, converging to $\lambda(\theta)$, is weakly-equicontinuous at every point.

Because of the uniform convergence of $\mathfrak{g}^n(\varphi)$ to $\mathfrak{g}(\varphi)$, the convergence of $\lambda^n(\theta)$ to $\lambda(\theta)$ leads to the convergence of $\mathfrak{g}^n(\lambda^n(\theta))$ to $\mathfrak{g}(\lambda(\theta))$. Likewise, the weak-equicontinuity of $\lambda^n(\theta)$ leads to the weak-equicontinuity of $\mathfrak{g}^n(\lambda^n(\theta))$.

Let $\mathfrak{p}^n$ be the potential surface associated with $\mathfrak{x}^n$, i.e., the potential surface with the same boundary values $\mathfrak{g}^n(\lambda^n(\theta))$ as $\mathfrak{x}^n$. It is a standard theorem of potential theory (see [10]) that $\mathfrak{p}^n$ converges to a potential surface $\mathfrak{p}$ whose boundary values are $\mathfrak{g}(\lambda(\theta))$ at the points of continuity of $\mathfrak{g}(\lambda(\theta))$. By (8.3) and (9.1),

$$D[\mathfrak{x}^n - \mathfrak{p}^n] \leqq \frac{k'}{k} A[\mathfrak{p}^n] \leqq \frac{k'K}{k} \cdot L(\Gamma^n)^2 \leqq M \tag{10.13}$$

since the lengths $L(\Gamma^n)$ are uniformly bounded. Letting $n \rightarrow \infty$,

(10.14) $$D[\mathfrak{x} - \mathfrak{p}] \leqq M.$$

Since $D[\mathfrak{x}]$ is finite by hypothesis of the theorem, it follows that $D[\mathfrak{p}]$ is finite.

The boundary values $\mathfrak{g}(\lambda(\theta))$ of $\mathfrak{p}$ have at most jump discontinuities. The finiteness of $D[\mathfrak{p}]$ then establishes (as in [12] and footnote 13) the continuity of $\mathfrak{p}$ in the closed unit circle. In particular, the boundary values of $\mathfrak{p}$ are continuous, so that at any point θ_0 of the boundary

$$\lim_{\theta \to \theta_0^+} \mathfrak{g}(\lambda(\theta)) = \lim_{\theta \to \theta_0^-} \mathfrak{g}(\lambda(\theta)).$$

This shows that any discontinuities that $\mathfrak{g}(\lambda(\theta))$ may have are removable discontinuities.

Because of the continuity of $\mathfrak{g}(\lambda(\theta))$, it follows that the sequence $\mathfrak{g}^n(\lambda^n(\theta))$ oscillates about the same value on both sides of the given point θ_0. That is, in the definition of weak-equicontinuity, the oscillation of $\mathfrak{g}^n(\lambda^n(\theta))$ on any closed (not necessarily connected) set Δ, which may lie on both sides of θ_0 but does not include θ_0, is less than ϵ for all $n \geqq N$. By using the Poisson integral representation for the potential function $\mathfrak{p}^n$, and this property of the sequence $\mathfrak{g}^n(\lambda^n(\theta))$, it is easy to demonstrate the weak equicontinuity of the sequence $\mathfrak{p}^n$.

Let θ_0 be any point on the unit circumference, and introduce polar coordinates (R, ψ) with the point θ_0 as pole. Indicate the region of the unit circle cut off by a circle of radius R about θ_0 by K_R, and the arc of this circle of radius R by K_R^*. Let ϵ be arbitrary, and let δ be the quantity stated in the definition of weak-equicontinuity for the sequence $\mathfrak{p}^n$, i.e., the diameter of each $\mathfrak{p}^n$ on a closed set in K_δ, not including θ_0, is less than ϵ for all sufficiently large n.

Let $\mathfrak{B}$ be any closed set in the region $K_{\eta\delta}$ not including the point θ_0, where $\eta = e^{-4\pi M^2/\epsilon^2}$ and M is the constant in (10.13). Because of (10.13), there is an arc $K_{\rho^n}^*$ where $\eta\delta \leqq \rho^n \leqq \delta$, on which the diameter of $\mathfrak{x}^n - \mathfrak{p}^n$ is less than ϵ.[13] Let δ_1 (depending on $\mathfrak{B}$) be such that $\mathfrak{B}$ lies outside K_{δ_1}. As previously, there is an arc $K_{\rho_1^n}^*$ where $\eta\delta_1 \leqq \rho_1^n \leqq \delta_1$ on which the diameter of $\mathfrak{x}^n - \mathfrak{p}^n$ is less than ϵ. Since $\mathfrak{x}^n - \mathfrak{p}^n = 0$ on the unit circumference, we see that the diameter of $\mathfrak{x}^n - \mathfrak{p}^n$ on the boundary of the ring shaped region enclosed by the arcs $K_{\rho_1^n}^*\ K_{\rho^n}^*$ is $< 2\epsilon$.

Now, by the weak-equicontinuity of $\mathfrak{p}^n$ and the choice of δ, the diameter of $\mathfrak{p}^n$ on the ring bounded by the arcs $K_{\eta\delta_1}^* K_\delta^*$ is less than ϵ for all sufficiently large n. For these n therefore, setting $\mathfrak{x}^n = (\mathfrak{x}^n - \mathfrak{p}^n) + \mathfrak{p}^n$, the diameter of $\mathfrak{x}^n$ on the boundary of the region enclosed by the arcs $K_{\rho_1^n}^* K_{\rho^n}^*$ is less than 3ϵ. By property 7.3 of an I-surface, the diameter of $\mathfrak{x}^n$ in the region $K_{\rho_1^n}^* K_{\rho^n}^*$ is less than $9c\epsilon$!! Because $\mathfrak{B}$ is included in this region, we have proved that the diameter of $\mathfrak{x}^n$ over the closed set $\mathfrak{B}$ is $< 9c\epsilon$ for all sufficiently large n.

Likewise, since the diameter of $\mathfrak{p}^n$ over $K_{\eta\delta_1}^* K_\delta^*$ is $< \epsilon$ for all sufficiently large n, the diameter of $\mathfrak{x}^n - \mathfrak{p}^n$ over $\mathfrak{B}$ is $< (9c + 1)\epsilon$.

[13] In any annular ring $\rho_1 \leqq r \leqq \rho_2$ there is an arc on which the diameter of $\mathfrak{p}$ is $\leqq \left(\dfrac{4\pi D[\mathfrak{p}]}{\log \rho_2/\rho_1}\right)^{1/2}$. See [12], p. 688–9.

By a passage to the limit it follows that the diameter of $\mathfrak{x}$ over the set $\mathfrak{B}$ is $\leqq 9c\epsilon$, while the diameter of $\mathfrak{x} - \mathfrak{p}$ over $\mathfrak{B}$ is $\leqq (9c + 1)\epsilon$. Since $\mathfrak{B}$ is any closed set lying in $K_{\eta\delta}$, but not including θ_0, the diameter of $\mathfrak{x}$ and $\mathfrak{x} - \mathfrak{p}$ over the region $K_{\eta\delta}$ is $\leqq 9c\epsilon$ and $(9c + 1)\epsilon$ respectively. This establishes the continuity of $\mathfrak{x}$ over the closed unit circle, and the identity of its boundary values with those of $\mathfrak{p}$. The theorem is proved.

Incidentally, if the finiteness of $D[\mathfrak{x}]$ is not assumed as it is in Theorem 10.4, one could still prove that $\mathfrak{x}$ behaves near the boundary exactly like the potential surface $\mathfrak{p}$.

Referring to the above theorem, if the limit curve is a Jordan curve only two situations can present themselves: the boundary values of $\mathfrak{x}$ are either constant or the whole of Γ. In the latter situation, the convergence of $\mathfrak{x}^n$ to $\mathfrak{x}$ is uniform in the closed unit circle.

Theorems 10.1 to 10.4 will now be combined into a form pertinent for later use.

THEOREM 10.5. *Let Γ^n be a sequence of curves of uniformly bounded length converging (in the Fréchet sense) to a curve Γ. Let $\mathfrak{x}^n$ be an I-surface bounded by Γ^n, $n = 1, 2, \cdots$, and suppose that $D[\mathfrak{x}^n] \leqq M$ for all n. Then a subsequence of the $\mathfrak{x}^n$ can be found which converges uniformly in every closed interior domain to an I-surface $\mathfrak{x}$ with boundary values lying monotonically on Γ.*

PROOF. This is a consequence of Theorems 10.1, 10.3, and 10.4.

11. The continuity theorem

As stated in the introduction, continuity considerations must play a vital role in the proof of the existence of unstable extremal surfaces. The following theorem is a generalization of a similar theorem due to Morse and Tompkins concerning the continuity of the area of potential surfaces ([20]).[14]

THEOREM 11.1. *Let $\mathfrak{x}^n$, bounded by Γ^n, be a sequence of I-surfaces converging uniformly to the I-surface $\mathfrak{x}$ bounded by Γ. Suppose that $L(\Gamma^n) \to L(\Gamma)$. Then*

$$J[\mathfrak{x}^n] \to J[\mathfrak{x}]$$

(and $A[\mathfrak{x}^n] \to A[\mathfrak{x}]$).

PROOF. We have already established, in Theorem 10.2, the continuity of all the relevant functionals when taken over any closed interior domain S. It remains to investigate a boundary strip, and to show that the area of the I-surfaces $\mathfrak{x}^n$ when taken over such a boundary strip can be made uniformly small by making the boundary strip sufficiently narrow. This will be done by using the analogous continuity theorem of Morse and Tompkins referring to potential surfaces.

Let ϵ be given. Because $D[\mathfrak{x}]$ exists there is a ρ' such that

$$D_{C_{\rho'1}}[\mathfrak{x}] < \epsilon.$$

[14] The continuity theorem and the isoperimetric inequality are closely related, as shown in [22].

By theorem 10.2,

$$D_{C_{\rho'\rho''}}[\mathfrak{x}^n - \mathfrak{x}] \to 0,$$

where $\rho'' = \dfrac{1 + \rho'}{2}$; by an easy estimation there is a ρ between ρ' and ρ'' such that the length of $\mathfrak{x}^n - \mathfrak{x}$ on C_ρ^* approaches 0. The length of $\mathfrak{x}^n$ on this circle C_ρ^* approaches the length of $\mathfrak{x}$ on C_ρ^*; and,

$$D_{C_{\rho 1}}[\mathfrak{x}] < \epsilon.$$

Let $\mathfrak{p}^n$, $\mathfrak{p}$ be the potential surfaces in the strip $C_{\rho 1}$ with the same values as $\mathfrak{x}^n$, $\mathfrak{x}$ on C_ρ and C_1. We have firstly

$$A_{C_{\rho 1}}[\mathfrak{p}] \leqq D_{C_{\rho 1}}[\mathfrak{p}] \leqq D_{C_{\rho 1}}[\mathfrak{x}] < \epsilon$$

and secondly by the theorem of Morse and Tompkins,

$$A_{C_{\rho 1}}[\mathfrak{p}^n] \to A_{C_{\rho 1}}[\mathfrak{p}]$$

since the lengths of the boundary curves are continuous. For sufficiently large n therefore,

$$A_{C_{\rho 1}}[\mathfrak{p}^n] < 2\epsilon$$

and

$$F_{C_{\rho 1}}[\mathfrak{p}^n] < 2k'\epsilon.$$

The connection between the I-surface $\mathfrak{x}^n$ and the potential surface $\mathfrak{p}^n$ is contained in Lemma 8.1. That lemma yields

$$F_{C_{\rho 1}}[\mathfrak{x}^n] \leqq F_{C_{\rho 1}}[\mathfrak{p}^n] < 2k'\epsilon,$$

so that, by 2.3,

$$A_{C_{\rho 1}}[\mathfrak{x}^n] < \frac{2k'}{m}\epsilon$$

and

$$J_{C_{\rho 1}}[\mathfrak{x}^n] < \left(2k' + \frac{2kk'}{m}\right)\epsilon. \tag{11.1}$$

A like inequality holds for $J_{C_{\rho 1}}[\mathfrak{x}]$.

These are the desired inequalities. They show that $A[\mathfrak{x}^n]$ and $J[\mathfrak{x}^n]$ can be made uniformly small when taken over a sufficiently narrow boundary strip. The completion of the proof is clear. By (11.1),

$$| J_{C_{\rho 1}}[\mathfrak{x}^n] - J_{C_{\rho 1}}[\mathfrak{x}] | < \left(4k' + \frac{4kk'}{m}\right)\epsilon. \tag{11.2}$$

And Theorem 10.2 (and Lemma 4.3) applied to the circle C_ρ gives

$$| J_{C_\rho}[\mathfrak{x}^n] - J_{C_\rho}[\mathfrak{x}] | < \epsilon$$

for sufficiently large n. Combining with (11.2),

$$|J[\mathfrak{x}^n] - J[\mathfrak{x}]| < \left(1 + 4k' + \frac{4kk'}{m}\right)\epsilon,$$

and Theorem 11.1 is proved.

IV. Extremal surfaces bounded by Γ

12. Extremal surfaces

We are now prepared to attack the main problem: to find surfaces bounded by the given rectifiable Jordan curve Γ which are critical for $I[\mathfrak{x}]$. The problem decomposes itself into two parts. In the first, the boundary values on the circumference of the (u, v) unit circle are held fixed—critical surfaces are therefore I-surfaces. In the second part, which is the concern of this section IV, the boundary values are varied. The variations which were used in the Plateau problem are suitable, and do not affect the integral $F[\mathfrak{x}]$. The results of these variations are identical to those in the Plateau problem, and we shall be able to conclude that $E = G$, $F = 0$ almost everywhere for a critical surface. This section IV is an elaboration of [22].

Before we proceed, we define an *extremal* surface. It is an I-surface for which $E = G$, $F = 0$ almost everywhere.[15] An extremal surface is extremal for the integral $J[\mathfrak{x}]$ as well, since (V_1) together with $E = G$, $F = 0$ almost everywhere can be written

$$\iint h_{\mathfrak{X}} \cdot \delta\mathfrak{X}\, du\, dv \geqq 0$$

where $h(\mathfrak{X}) = f(\mathfrak{X}) + k\,|\,\mathfrak{X}\,|$.

For an extremal surface $\mathfrak{x}$, $J[\mathfrak{x}] = I[\mathfrak{x}]$. Every theorem in section III yields a corresponding theorem involving the I-integral of an extremal surface. The following are pertinent.

THEOREM 12.1. *If $\mathfrak{x}$ is an extremal surface bounded by a curve of length L, then*

$$I[\mathfrak{x}] \leqq \kappa L^2,$$

where κ is a constant.

THEOREM 12.2. *Let $\mathfrak{x}^n$ be a sequence of extremal surfaces bounded by Γ^n converging to the surface $\mathfrak{x}$ bounded by Γ, and suppose that $L(\Gamma^n) \to L(\Gamma)$. Then*

(a) $I[\mathfrak{x}^n] \to I[\mathfrak{x}]$;

(b) *$\mathfrak{x}$ is an extremal surface.*

PROOFS. Theorem 12.1 is an immediate consequence of Theorem 9.1. Concerning Theorem 12.2, we have

---

[15] There is an interesting question related to this. Can a given surface have two different representations of class $\mathfrak{T}$, in both of which $E = G$, $F = 0$ almost everywhere, without being related by a linear transformation of the unit circle? Counter-examples are easily constructed if the representations are not required to be of class $\mathfrak{T}$.

(12.1) $$J[\mathfrak{x}] = \lim J[\mathfrak{x}^n] = \lim I[\mathfrak{x}^n]$$

by Theorem 11.1, and

(12.2) $$I[\mathfrak{x}] \leqq \lim\inf I[\mathfrak{x}^n].$$

The inequality $J[\mathfrak{x}] \leqq I[\mathfrak{x}]$ and (12.1), (12.2) give

$$I[\mathfrak{x}] = \lim I[\mathfrak{x}^n]$$

and

$$I[\mathfrak{x}] = J[\mathfrak{x}].$$

The first of these equations is Theorem 12.2a), while the second of them establishes $D[\mathfrak{x}] = A[\mathfrak{x}]$ which can occur if and only if $E = G, F = 0$ almost everywhere. This proves Theorem 12.2b) since $\mathfrak{x}$ is already known to be an I-surface by Theorem 10.3.

13. The space $\mathfrak{A}$ of admitted surfaces and the approximating spaces

DEFINITION 13.1. The space $\mathfrak{A}$ consists of surfaces $\mathfrak{x}$ satisfying:
(i) $\mathfrak{x}$ is of class $\mathfrak{T}$;
(ii) $\mathfrak{x}$ maps the unit circumference monotonically onto Γ;
(iii) $\mathfrak{x}$ maps three specified points $\theta_1, \theta_2, \theta_3$ of the unit circumference into three specified points P_1, P_2, P_3 of Γ (three-point condition).
The distance $|\mathfrak{x} - \mathfrak{y}|_1$ between two surfaces of $\mathfrak{A}$ is defined as

(13.1) $$|\mathfrak{x} - \mathfrak{y}|_1 = \sum_{n=1}^{\infty} \frac{1}{2^n} \underset{r \leqq 1-(1/2n)}{\text{maximum}} |\mathfrak{x}(r, \theta) - \mathfrak{y}(r, \theta)|$$

using polar coordinates (r, θ).

For given boundary values, $I[\mathfrak{x}]$ is minimized by an I-surface. We therefore introduce a subspace $\mathfrak{J}$ of $\mathfrak{A}$.

DEFINITION 13.2. $\mathfrak{J}$ is the subspace of $\mathfrak{A}$ consisting of the I-surfaces in $\mathfrak{A}$. (By the comment before Theorem 10.5, the metric (13.1) is equivalent to the uniform metric in $\mathfrak{J}$.)

By Theorems 6.1 and 5.3, the space $\mathfrak{A}$ can be I-retracted into the subspace $\mathfrak{J}$, so that our considerations may be limited to $\mathfrak{J}$. And $\mathfrak{J}$ has the following property, an immediate consequence of Theorem 10.5.

THEOREM 13.1. *$\mathfrak{J}_N$ is compact, where $\mathfrak{J}_N$ consists of those surfaces $\mathfrak{x}$ in $\mathfrak{J}$ for which $I[\mathfrak{x}] \leqq N$.*

A descriptive name for a subspace of $\mathfrak{A}$ which enjoys the retraction property stated above and the compactness property of Theorem 13.1 is a *'core'* of $\mathfrak{A}$.

To establish the existence of unstable critical points it is necessary to prove some attribute of the functional $I[\mathfrak{x}]$ and the space $\mathfrak{J}$ analogous to 'reducibility'. This is usually difficult and has been accomplished in the Plateau problem for specific classes of boundary curves (cf. [17], [18], [19]). This obstacle is overcome here by introducing certain approximating spaces.

The approximating spaces $\mathfrak{A}^H$ will be obtained by considering curves lying in

certain piecewise convex regions surrounding Γ. Decompose Γ into m consecutive arcs $\alpha_1, \alpha_2, \cdots, \alpha_m$ (denote the collection $\{\alpha_1, \alpha_2, \cdots, \alpha_m\}$ by α), and enclose each α_i in a closed convex point set H_i (e.g., by forming the convex hull of α_i). The piecewise convex region is the point set $H = \sum_{i=1}^{m} H_i$. Such a covering of Γ by convex sets H_i will be called a *convex covering* of Γ. The maximum of the diameters of H_i, $i = 1, 2, \cdots, m$, is the diameter of the covering and is denoted by $\delta(H)$.

It is convenient to limit the present considerations to a special kind of rectifiable curve Γ, a Jordan curve *of type* $\mathfrak{K}$. It is a curve which admits of convex coverings of arbitrarily small diameter with the property: *only consecutive H_i's have a non-zero intersection.* In section 17 we shall sketch the procedure in the general case.

DEFINITION 13.3. The elements of the space $\mathfrak{A}^H$ are surfaces $\mathfrak{x}$ satisfying the following conditions.

(i) $\mathfrak{x}$ is of class $\mathfrak{T}$.

(ii) The boundary of $\mathfrak{x}$ has length $\leqq L(\Gamma)$.

(iii) The unit circumference can be decomposed into m consecutive overlapping arcs $\beta_1, \beta_2, \cdots, \beta_m$ which are mapped by $\mathfrak{x}$ into $H_1, H_2, \cdots, H_m$ respectively (the set $\{\beta_1, \beta_2, \cdots, \beta_m\}$ is denoted collectively by β).

(iiii) There are arcs β_i containing θ_1, θ_2, or θ_3 which correspond to sets H_i containing P_1, P_2, or P_3 respectively (three point condition).

Also, surfaces which differ by a constant are identified. The distance $|\mathfrak{x} - \mathfrak{y}|_2$ is defined by

$$|\mathfrak{x} - \mathfrak{y}|_2 = \liminf_{\mathfrak{c}} |\mathfrak{x} - \mathfrak{y} + \mathfrak{c}|_1$$

where $|\cdots|_1$ is defined in (13.1).

If $\mathfrak{x}$ and $\mathfrak{y}$ differ by a constant, $I[\mathfrak{x}] = I[\mathfrak{y}]$. The totality of all surfaces identified to a given surface will still be called a 'surface'.

It is clear that $\mathfrak{A}^H \supset \mathfrak{A}$.

DEFINITION 13.4. The subset of $\mathfrak{A}^H$ consisting of those surfaces which satisfy definition 16.1c) for a given set $\beta = \{\beta_1, \beta_2, \cdots, \beta_m\}$ of arcs is denoted by $\mathfrak{A}^H(\beta)$.

The significant properties of $\mathfrak{A}^H(\beta)$ are:

(a) $\mathfrak{A}^H(\beta)$ *is convex*—if $\mathfrak{x}_0$ and $\mathfrak{x}_1$ belong to $\mathfrak{A}^H(\beta)$, so does $\mathfrak{x}_t = (1 - t)\mathfrak{x}_0 + t\mathfrak{x}_1$ for $0 \leqq t \leqq 1$. This follows from the convexity of each H_i and from the convexity of the length of the boundary of $\mathfrak{x}_t$ as a function of t: $L[\mathfrak{x}_t] \leqq (1 - t)L[\mathfrak{x}_0] + tL[\mathfrak{x}_1]$.

(b) $\mathfrak{A}^H(\beta)$ *is closed*—if the surface $\mathfrak{x}$ of class $\mathfrak{T}$ is the *uniform* limit of a sequence of surfaces in $\mathfrak{A}^H(\beta)$ then $\mathfrak{x}$ lies in $\mathfrak{A}^H(\beta)$. This follows from the closedness of each point *set H_i and the lower* semicontinuity of length.

(c) If $\mathfrak{x}$ and $\mathfrak{y}$ are any two surfaces in $\mathfrak{A}^H(\beta)$, the boundary values of $\mathfrak{x} - \mathfrak{y}$ are in absolute value $\leqq 2\delta(H)$.

The properties a) and b) permit the application of the theory of convex sets

of surfaces developed in II, §§5, 6. Consider the problem of minimizing $I[\mathfrak{x}]$ in the set $\mathfrak{A}^H(\beta)$. By Theorems 6.1, 6.2, it suffices to consider for this minimum problem merely the I-surfaces in $\mathfrak{A}^H(\beta)$. Let $\mathfrak{x}^n$ be a minimizing sequence of I-surfaces. By Theorem 10.5, there is a limit surface $\mathfrak{x}$ which is an I-surface in $\mathfrak{A}^H(\beta)$—this is the solution to the minimum problem. By Theorem 5.1 the solution is unique (except for translations, but all translated surfaces have been identified). The solution is denoted by $\mathfrak{x}(\beta)$.

DEFINITION 13.5. The space $\mathfrak{Z}^H$ is the subspace of $\mathfrak{A}^H$ consisting of the surfaces $\mathfrak{x}(\beta)$ for all possible β.

THEOREM 13.2.[16] *The surface $\mathfrak{x}(\beta)$ and the quantity $I[\mathfrak{x}(\beta)]$ depend continuously on the collection $\beta = \{\beta_1, \beta_2, \cdots, \beta_m\}$.*

PROOF. Let β^n be a sequence approaching β. Abbreviate the surface $\mathfrak{x}(\beta)$ by $\mathfrak{x}$. Perform the following transformation of the unit circle into itself:[17]

$$\begin{cases} R = re^{\epsilon\lambda(r,\theta)} \\ \phi = \theta + \epsilon\mu(r, \theta), \end{cases} \qquad \lambda(1, \theta) = 0 \tag{13.2}$$

where $\lambda(r, \theta)$, $\mu(r, \theta)$ have continuous and bounded first derivatives,

$$\{|\lambda_r|, |\lambda_\theta|, |\mu_r|, |\mu_\theta|\} < M$$

and where $|\epsilon| \leqq \dfrac{1}{4M}$. The Jacobean of the transformation is $\geqq \frac{1}{4}e^{\epsilon\lambda(r,\theta)} > 0$. Vary $\mathfrak{x}$ by setting

$$\mathfrak{y}_\epsilon(R, \phi) = \mathfrak{x}(r, \theta). \tag{13.3}$$

The functional $F[\mathfrak{x}]$ is invariant under changes of parameters,

$$F[\mathfrak{y}_\epsilon] = F[\mathfrak{x}]. \tag{13.4}$$

Concerning the Dirichlet integral,

$$D[\mathfrak{y}_\epsilon] = D[\mathfrak{x}] + \frac{\epsilon}{2}\iint \{(\mu_\theta - r\lambda_r)(E' - G') - (r\mu_r + \lambda_\theta)2F'\} r\,dr\,d\theta + \epsilon^2 B \tag{13.5}$$

where $E' = \mathfrak{x}_r^2$, $F' = \dfrac{1}{r}\mathfrak{x}_r\mathfrak{x}_\theta$, $G' = \dfrac{1}{r^2}\mathfrak{x}_\theta^2$, and B is a certain integral which can be estimated by

$$|B| \leqq 144D[\mathfrak{x}]\cdot M^2. \tag{13.6}$$

Thus, $\mathfrak{y}_\epsilon$ is of class $\mathfrak{T}$ and belongs to $\mathfrak{A}^H$. And,

$$I[\mathfrak{y}_\epsilon] \to I[\mathfrak{x}] \quad \text{if} \quad \epsilon M \to 0. \tag{13.7}$$

Setting $\lambda^n(r, \theta) \equiv 0$, it is elementary to select $\epsilon^n\mu^n(r, \theta)$, so that the image of β under the transformation (13.2) is β^n and such that the first derivatives of $\epsilon^n\mu^n(r, \theta)$ uniformly approach 0 as $n \to \infty$. The surface $\mathfrak{y}_{\epsilon^n}^n$ then belongs to $\mathfrak{A}^H(\beta^n)$, and (13.7) yields

[16] This theorem and its proof are modelled after [19].

[17] For similar transformations and detailed calculations, see [17], p. 850.

$$I[\mathfrak{x}(\beta)] = \lim I[\mathfrak{y}^n_{\epsilon^n}] \geqq \limsup I[\mathfrak{x}(\beta^n)]. \tag{13.8}$$

In particular, the quantities $I[\mathfrak{x}(\beta^n)]$ are uniformly bounded.

On the other hand, by Theorem 10.5 a subsequence of the $\mathfrak{x}(\beta^n)$ converges to a surface $\mathfrak{z}$ in $\mathfrak{A}^H(\beta)$. Hence,

$$I[\mathfrak{x}(\beta)] \leqq I[\mathfrak{z}] \leqq \liminf I[\mathfrak{x}(\beta^n)]. \tag{13.9}$$

Inequalities (13.8) and (13.9) result in

$$I[\mathfrak{x}(\beta)] = \lim I[\mathfrak{x}(\beta^n)] \tag{13.10}$$

and

$$I[\mathfrak{z}] = I[\mathfrak{x}(\beta)].$$

But then $\mathfrak{z} = \mathfrak{x}(\beta)$, and

$$\mathfrak{x}(\beta^n) \rightarrow \mathfrak{x}(\beta). \tag{13.11}$$

Relations (13.10) and (13.11) are the statements of the theorem.

Theorem 13.2 can be stated in a more revealing form. It is equivalent to the following two theorems.

THEOREM 13.3. *$\mathfrak{J}^H_N$ is compact.*

THEOREM 13.4. *The functional $I[\mathfrak{x}]$ is continuous in the space $\mathfrak{J}^H$.*

Also, by Theorem 5.3, $\mathfrak{A}^H$ can be I-retracted into $\mathfrak{J}^H$. The subspace $\mathfrak{J}^H$ is a *core* of $\mathfrak{A}^H$.

14. Unstable critical points in $\mathfrak{J}^H$.

Theorem 13.4 shows that the general critical point theory of Morse applies to the functional $I[\mathfrak{x}]$ in the space $\mathfrak{J}^H$. However, for the limited purpose of demonstrating the existence of unstable extremals whenever two stable extremals exist, it is not necessary to resort to the Morse theory. One need merely proceed as follows. Given two surfaces $\mathfrak{x}$ and $\mathfrak{x}'$ in $\mathfrak{J}^H$, join them by a *minimizing* closed connected set C of surfaces in $\mathfrak{J}^H$, i.e., a closed connected set C on which the maximum i of $I[\mathfrak{x}]$ is the smallest possible among all closed connected sets in $\mathfrak{J}^H$ joining $\mathfrak{x}$, $\mathfrak{x}'$. Such a minimizing connected set C exists as a result of theorems 13.3 and 13.4. If i is greater than both $I[\mathfrak{x}]$ and $I[\mathfrak{x}']$, it is easy to prove (by a Heine-Borel argument) that C must contain a surface $\mathfrak{z}$ at its 'top', $I[\mathfrak{z}] = i$, which is '*critical*'. A surface $\mathfrak{z}$, with $I[\mathfrak{z}] = i$, is said to be '*critical*' if no neighborhood of it in $\mathfrak{J}^H_i$ can be deformed in $\mathfrak{J}^H_i$ into a set lying in $\mathfrak{J}^H_{i-\eta}$ for some positive η. Criticalness is a topologic attribute of a surface, while extremality is analytic. We shall show in the next section that a critical surface is extremal (but not necessarily the reverse). Anticipating this result, we obtain from Theorem 12.1 that $I[\mathfrak{z}] \leqq \kappa L^2$ for any critical surface, where L is the length of Γ.

Summarizing, we may state the following theorem.

THEOREM 14.1. *Any two surfaces $\mathfrak{x}_1$, $\mathfrak{x}_2$ in $\mathfrak{J}^H$ can be joined by a minimizing closed connected set C of surfaces in $\mathfrak{J}^H$. The maximum i of $I[\mathfrak{x}]$ on C is* $\leqq$ max.

$\{I[\mathfrak{x}_1], I[\mathfrak{x}_2], \kappa L^2\}$; *if* $i >$ max. $\{I[\mathfrak{x}_1], I[\mathfrak{x}_2]\}$ *then C contains an extremal surface* $\mathfrak{z}$ *for which* $I[\mathfrak{z}] = i$.

15. Critical surfaces are extremal surfaces

For any surface $\mathfrak{x} = \mathfrak{x}(\beta)$ in $\mathfrak{I}^H$ denote the coefficient of ϵ in the expansion of $I[\mathfrak{y}_\epsilon]$ due to the variation (13.2), (13.3) by $\delta I[\mathfrak{x}]$. By (13.4), (13.5),

$$\delta I[\mathfrak{x}] = \frac{k}{2}\iint \{(\mu_\theta - r\lambda_r)(E' - G') - (r\mu_r + \lambda_\theta)2F'\}r\,dr\,d\theta. \tag{15.1}$$

LEMMA 15.1. *For given* λ, μ, $\delta I[\mathfrak{x}]$ *depends continuously on* $\mathfrak{x}$ *in* $\mathfrak{I}^H$.

PROOF. The lemma is a consequence of the continuity of $I[\mathfrak{x}]$, and therefore of $D[\mathfrak{x}]$, in $\mathfrak{I}^H$ and Lemmas 4.1, 4.2.

Suppose that for a surface $\mathfrak{x}_0$ in $\mathfrak{I}^H$ we had $\delta I[\mathfrak{x}_0] \neq 0$ for some choice of λ, μ. By Lemma 15.1 there is a neighborhood of $\mathfrak{x}_0$ in $\mathfrak{I}^H$ such that $\delta I[\mathfrak{x}]$ has the same sign as $\delta I[\mathfrak{x}_0]$ and

$$|\,\delta I[\mathfrak{x}]\,| > \tfrac{1}{2}\,|\,\delta I[\mathfrak{x}_0]\,| \tag{15.2}$$

for any surface $\mathfrak{x}$ in this neighborhood. Let β^ϵ be the image of β under the transformation (13.2), and consider the surface $\mathfrak{x}(\beta^\epsilon)$ in $\mathfrak{I}^H$. From (13.4) and (13.5),

$$I[\mathfrak{x}(\beta^\epsilon)] \leqq I[\mathfrak{y}_\epsilon] = I[\mathfrak{x}] + \epsilon\delta I[\mathfrak{x}] + \epsilon^2 kB.$$

Selecting the sign of ϵ opposite to that of $\delta I[\mathfrak{x}_0]$, limiting ϵ to $0 \leqq |\,\epsilon\,| \leqq \dfrac{1}{144kM^2D[\mathfrak{x}]}\cdot\dfrac{|\,\delta I[\mathfrak{x}_0]\,|}{4}$, and using (15.2) and (13.6), we have

$$I[\mathfrak{x}(\beta^\epsilon)] \leqq I[\mathfrak{x}] - \frac{|\,\epsilon\,|}{4}\,|\,\delta I[\mathfrak{x}_0]\,|.$$

This inequality shows that a neighborhood of $\mathfrak{x}_0$ can be I-deformed so as to decrease the I-integral by at least the amount $\dfrac{|\,\delta I[\mathfrak{x}_0]\,|^2}{144kM^2D[\mathfrak{x}]\cdot 4^2}$. In other words, $\mathfrak{x}_0$ is not a critical surface. We have proved the next Theorem 15.1.

THEOREM 15.1. *A critical surface in* $\mathfrak{I}^H$ *must satisfy*

$$\iint \{(\mu_\theta - r\lambda_r)(E' - G') - (r\mu_r + \lambda_\theta)2F'\}r\,dr\,d\theta = 0 \tag{V_2}$$

for all λ, μ *with continuous and bounded first derivatives satisfying* $\lambda(1, \theta) = 0$.

THEOREM 15.2. *A critical surface in* $\mathfrak{I}^H$ *is an extremal surface.*

PROOF. It is required to show that (V_2) implies $E' = G'$, $F' = 0$ almost everywhere. Let $P(u, v)$, $Q(u, v)$ be any two polynomials and determine λ, μ so that

$$\lambda(1, \theta) = 0 \tag{15.3}$$

and

$$(15.4) \qquad \begin{cases} \mu_\theta - r\lambda_r = r^2 P \\ r\mu_r + \lambda_\theta = -r^2 Q. \end{cases}$$

This is possible because one can first choose λ to satisfy (15.3) and

$$(15.5) \qquad \Delta\lambda = -(Q_\theta + 2P + rP_r)$$

where Δ is the potential operator. It is well known that such a function λ exists (and in fact can be expressed in the form of an integral) and has continuous and bounded first derivatives (and is even analytic including the unit circumference). Then μ can be determined from (15.4). For this choice of λ, μ, (V_2) becomes

$$(15.6) \qquad \iint \{P(E' - G') + QF'\} r^3 \, dr \, d\theta = 0$$

for any two polynomials P, Q.

It follows by a passage to the limit that (15.6) also holds for any two bounded measurable functions P, Q.

Select

$$\begin{cases} P = \text{signum } (E' - G') \\ Q = \text{signum } F' \end{cases}$$

where signum $f = +1, -1$, or 0 according as $f > 0, f < 0$, or $f = 0$ or undefined respectively. Equation (15.6) becomes

$$\iint \{| E' - G' | + | F' |\} r^3 \, dr \, d\theta = 0$$

or $E' - G' = 0$, $F' = 0$ almost everywhere. Q.E.D.

16. Passage to the limit to obtain unstable extremal surfaces in $\mathfrak{J}$

The passage to the limit will be conducted by selecting a sequence of piecewise convex regions H surrounding Γ for which $\delta(H) \to 0$. Every surface in $\mathfrak{A}^H$ has a boundary curve which approaches Γ in the Fréchet sense as $\delta(H) \to 0$. We expect that $\mathfrak{J}^H$ approaches $\mathfrak{J}$ as $\delta(H) \to 0$, and this is the content of the next theorem.

THEOREM 16.1. $\mathfrak{J}_N = \lim \mathfrak{J}_N^H$ as $\delta(H) \to 0$.

PROOF. It is clear from Theorem 10.5 that every limit surface for the sequence $\mathfrak{J}_N^H$ is a surface in $\mathfrak{J}_N$, i.e.,

$$\mathfrak{J}_N \supset \lim \mathfrak{J}_N^H.$$

For an inequality in the reversed sense, let $\mathfrak{x}$ be any surface in $\mathfrak{J}_N$, and determine for each H a β^H such that $\mathfrak{x}$ belongs to $\mathfrak{A}^H(\beta^H)$. Consider the surface $\mathfrak{x}^H(\beta^H)$ in $\mathfrak{J}^H$. We have $I[\mathfrak{x}^H(\beta^H)] \leq I[\mathfrak{x}] \leq N$, and the boundary values of $\mathfrak{x}^H(\beta^H)$ differ from those of $\mathfrak{x}$ in absolute value by at most $2 \cdot \delta(H)$. Letting $\delta(H) \to 0$, the boundary values of $\mathfrak{x}^H(\beta^H)$ converge uniformly to the boundary values of $\mathfrak{x}$, and so $\mathfrak{x}^H(\beta^H)$ converges uniformly to an I-surface $\mathfrak{y}$ with the same

boundary values as $\mathfrak{x}$. But then $\mathfrak{y} \equiv \mathfrak{x}$. Thus, $\mathfrak{x}^H(\beta^H)$ converges uniformly to $\mathfrak{x}$ as $\delta(H) \to 0$, and $\mathfrak{J}_N \subset \lim \mathfrak{J}_N^H$.

MAIN THEOREM 16.2. *If the rectifiable Jordan curve Γ of type $\mathfrak{K}$ (defined in §13, p. 568) bounds two extremal surfaces which are proper relative minima, then Γ bounds an unstable extremal surface.*

PROOF. Let the two extremal surfaces be $\mathfrak{x}'$, $\mathfrak{x}''$. Let H be any piecewise convex region surrounding Γ. Suppose that $\mathfrak{x}' \subset \mathfrak{A}^H(\beta')$, $\mathfrak{x}'' \subset \mathfrak{A}^H(\beta'')$ and consider $\mathfrak{x}^H(\beta')$, $\mathfrak{x}^H(\beta'')$ in $\mathfrak{J}^H$. By Theorem 14.1, $\mathfrak{x}^H(\beta')$, $\mathfrak{x}^H(\beta'')$ can be joined by a minimizing connected set C^H lying in $\mathfrak{J}^H_{i^H}$ where i^H is the maximum of $I[\mathfrak{x}]$ on C^H and $i^H \leqq \max. \{I[\mathfrak{x}'], I[\mathfrak{x}''], \kappa L^2\}$.

Choosing a sequence of H's for which $\delta(H) \to 0$ and for which i^H approaches a limit i, all the limiting surfaces of the sets C^H form a *connected* set C in I_i. This follows from Theorems 16.1 and 13.3. The connected set C contains $\mathfrak{x}'$ and $\mathfrak{x}''$. Since $\mathfrak{x}'$ and $\mathfrak{x}''$ were *proper* relative minima, $i > \max. \{I[\mathfrak{x}'], I[\mathfrak{x}'']\}$. Hence for all H with sufficiently small $\delta(H)$,

$$i^H > \max. \{I[\mathfrak{x}'], I[\mathfrak{x}'']\} \geqq \max. \{I[\mathfrak{x}^H(\beta')], I[\mathfrak{x}^H(\beta'')]\}.$$

Theorem 14.1 shows that there must be an extremal surface $\mathfrak{z}^H$ on C^H for which $I[\mathfrak{z}^H] = i^H$.

A subsequence of the $\mathfrak{z}^H$'s approaches a surface $\mathfrak{z}$ which is extremal by theorem 12.2(b), and for which $I[\mathfrak{z}] = i$ by Theorem 12.2(a). This extremal surface $\mathfrak{z}$ lies on C and is the required unstable extremal surface. Q.E.D.

17. A sketch of the procedure for any rectifiable Jordan curve Γ

We have heretofore supposed that Γ was a Jordan curve of type $\mathfrak{K}$. If this limitation is removed certain modifications must be made, beginning with the Definition 13.3 of $\mathfrak{A}^H$. The surfaces admitted to $\mathfrak{A}^H$ must now include *degenerate* surfaces composed of several pieces. Degenerate surfaces of a different type but with a similar motivation were introduced in [21], and we refer the reader to [21] for a discussion more adequate than the brief description below.

A degenerate surface $\mathfrak{x}$ is defined over several unit circles, each of the various pieces being called a *constituent* of $\mathfrak{x}$. The unit circumference for each constituent is divided successively into arcs $\beta_{i_1}, \beta_{i_2}, \cdots, \beta_{i_k}$ where $i_1 < i_2 < \cdots < i_k$ and the corresponding $H_{i_1}, H_{i_2}, \cdots, H_{i_k}$ consecutively overlap. The various constituents are interrelated as follows: it is possible to join the unit circumference at single points where $\mathfrak{x}$ has the same value, breaking the circumferences at these points, so that the arcs $\beta_1, \beta_2, \cdots, \beta_m$ arrange themselves consecutively. Finally, each constituent of $\mathfrak{x}$ is normalized, e.g., by a suitable three point condition.

A sequence of ordinary, non-degenerate, surfaces can converge to a degenerate surface only if some of the arcs β_i approach a point. The limit degenerate surface will consist of the limit over the unit circle in the usual sense plus all other limits over different unit circles obtained by performing suitable linear transformations on the surfaces of the sequence.

Continuing with the development of section 13, one has the subset $\mathfrak{A}''(\beta)$ and the surface $\mathfrak{x}(\beta)$, where β now includes the type of domain as well as the positions of the arcs $\beta_1, \cdots, \beta_m$. To prove theorem 13.2 in case the limit surface $\mathfrak{x}$ is degenerate, vary each of the constituents of $\mathfrak{x}$ as in §13 and then piece them together by using a device similar to one in [13], p. 79, 80. This would establish the continuity of $I[\mathfrak{x}(\beta)]$ and with it Theorems 13.2, 13.3, 13.4.

A degenerate surface is called extremal if each of its constituents is extremal. The isoperimetric inequality still applies since it applies to each of the constituents and the total length of all the boundaries is $\leqq L$. Theorem 15.2 is the only remaining fact required for the proof of Theorem 14.1.

To establish Theorem 15.2, let $\mathfrak{x}_0$ be a degenerate surface of which one of the constituents $\mathfrak{x}_0'$ is not extremal. The expression (15.6), taken over the unit circle C' of $\mathfrak{x}_0'$, is $\neq 0$ for some choice P, Q of functions which vanish in a strip near the boundary of C' and have continuous derivatives. Select λ, μ to satisfy (15.3) and (15.4). Normalize all surfaces $\mathfrak{x}$ near $\mathfrak{x}_0$ so that one of the constituents of $\mathfrak{x}$ is normalized in the same manner as $\mathfrak{x}_0'$. The expression (15.1) for $I[\mathfrak{x}]$ varies continuously with $\mathfrak{x}$ since $\mu_\theta - r\lambda_r = r\mu_r + \lambda_\theta = 0$ in a boundary strip and $I[\mathfrak{x}]$ is continuous in closed interior domains (theorem 10.2). The reasoning of §15 completes Theorem 15.2.

Lastly, §16 can be repaired as follows. If $\mathfrak{x}$ is degenerate, define the *major* constituent of $\mathfrak{x}$ as that constituent whose boundary curve is the longest (if $\delta(H)$ is small). Denote the major constituent by $\mathfrak{x}'$, its boundary curve by Γ', the length of Γ' by L'. Let the lengths of the boundaries of all the other constituents of $\mathfrak{x}$ be $L_2, L_3, \cdots, L_k$, so that $L' + L_2 + L_3 + \cdots + L_k \leqq L$. If $\delta(H) \to 0$, it is clear since Γ is a Jordan curve that $\Gamma' \to \Gamma$, whence $L' \to L$. Thus all the other lengths $L_2, L_3, \cdots \to 0$, and the limit of $\mathfrak{x}$ is merely the limit of its major constituent in the usual sense. Also, the value of the integral I over all the constituents other than $\mathfrak{x}'$ is $\leqq \kappa(L_2^2 + L_3^2 + \cdots) \leqq \kappa(L - L')^2$, which approaches 0 as $\delta(H) \to 0$. Hence $I[\mathfrak{x}] - I[\mathfrak{x}'] \to 0$ as $\delta(H) \to 0$, and the continuity Theorem 12.2 applies. The main Theorem 16.2 is established for any rectifiable Jordan curve Γ.

Remarks. We shall first comment on the restrictions made in §2 concerning the integrand of our double integral problem. The first restriction is that the integrand does not involve x, y, z explicitly, an assumption used rather essentially in §6 where a 'limit' surface is obtained from a minimizing sequence. If this limitation is removed, the 'limit' surface can be obtained in some other manner, e.g., in the sense of weak convergence as in the existence theory of Morrey [5]. The corresponding modifications, beginning with the variational condition in Theorems 5.2 and 6.2, would be unessential.

The other major assumption is 2.4, which asserts that the integral under consideration is dominantly an area integral. This drastic limitiation was used in a vital way in the uniqueness theorem of §5, and it is not so easily removed. Without this assumption, the uniqueness Theorem 5.1 may fail, causing the whole structure of this paper to collapse. One could still discuss I-surfaces, and many

of the theorems of part III would be valid. But it is of no avail to introduce the space $\mathfrak{Z}^H$. It seems to the author that it might suffice to establish the validity of the Morse theory for the set $\mathfrak{B}$ of surfaces with given boundary values.

Concerning the general Morse relations of which our main Theorem 16.2 is a special case, it should be pointed out that it is possible to establish a slightly weaker form of the usual Morse theory (cf. [22]). But we shall not prove this result here since it indicates that the usual Morse theory is in need of revision. Furthermore, the following question is still unanswered: how to characterize an extremal surface of given type.[18] In the usual Morse theory, this problem is completely divorced from all other considerations.

Finally, let us note that the case of polygonal boundaries does not require any passage to the limit (cf. [19]). And one can establish the existence of unstable extremal surfaces bounded by several non-intersecting rectifiable Jordan curves by using the method in [21].

College of the City of New York.

Bibliography

a. Multiple integral problems

[1] L. Tonelli, *Sur la semi-continuité des intégrales doubles du calcul des variations*, Acta Math., 53, 325–346 (1929).

[2] L. Tonelli, *L'estremo assoluto degli integrali doppi*, Ann. Scuola Norm. Sup. Pisa, 2, 89–130 (1933).

[3] E. J. McShane, *Semi-continuity of double integral problems in the calculus of variations*, Ann. Math., 33, 460–484 (1932).

[4] E. J. McShane, *Existence theorems for double integral problems of the calculus of variations*, Trans. Amer. Math. Soc., 38, 549–563 (1935).

[5] C. B. Morrey, *Existence and differentiability theorems for the solutions of variational problems for multiple integrals*, Bull. Am. Math. Soc., 46, 439–458 (1940).

[6] T. Radó, *On the semicontinuity of double integrals in parametric form*, Trans. Am. Math. Soc., 51, 336–361 (1942).

b. Plateau problem

[7] J. Douglas, *Solution of the problem of Plateau*, Trans. Am. Math. Soc., 33, 263–321 (1931).

[8] T. Radó, *On Plateau's problem*, Annals Math., 31, 457–469 (1930).

[9] J. Douglas, *The problem of Plateau*, Bull. Am. Math. Soc., 227–251, (1933).

[10] T. Radó, *On the problem of Plateau*, Ergeb. Math., II, 2 (1933).

[11] E. J. McShane, *Parametrizations of saddle surfaces with applications to the problem of Plateau*, Trans. Am. Math. Soc., 35, 716–733, (1933).

[12] R. Courant, *Plateau's problem and Dirichlet's principle*, Ann. Math., 38, 679–724 (1937).

[13] R. Courant, *The existence of minimal surfaces of given topological structure under prescribed boundary conditions*, Acta Math., 72, 51–98, (1940).

c. General critical point theory

[14] M. Morse, *The calculus of variations in the large*, Amer. Math. Soc. Coll. Publ., vol. 18, 1934.

[18] For the case of minimal surfaces, reference can be made to the classical work of H. A. Schwarz, *Collected works*, I, pp. 151–167, 223–269.

[15] M. Morse, *Analysis in the large*, notes of Institute for Advanced Study, 1936-37.

[16] M. Morse, *Functional topology and abstract variational theory*, Ann. Math., 38, 386-448, (1937).

d. Unstable minimal surfaces

[17] M. Shiffman, *The Plateau problem for non-relative minima*, Ann. Math., 40, 834-854 (1939).

[18] M. Morse and C. Tompkins, *The existence of minimal surfaces of general critical type*, Ann. Math., 40, 443-472 (1939).

[19] R. Courant, *Critical points and unstable minimal surfaces*, Proc. Nat. Acad. Sci., 27, 51-57 (1941).

[20] M. Morse and C. Tompkins, *The continuity of the area of harmonic surfaces as a function of the boundary representation*, Amer. Jour. Math., 63, 825-838, (1941).

[21] M. Shiffman, *Unstable minimal surfaces with several boundaries*, Ann. Math., 43, 197-222, (1942).

[22] M. Shiffman, *Unstable minimal surfaces with any rectifiable boundary*, Proc. Nat. Acad. Sci., 28, 103-108 (1942).

ANNALS OF MATHEMATICS
Vol. 45, No. 3, July, 1944

ON THE THEORY OF INDEFINITE QUADRATIC FORMS

BY CARL LUDWIG SIEGEL

(Received December 20, 1943)

1. Introduction

Let $\mathfrak{S}$ be a non-singular real m-rowed symmetric matrix and let $\Omega(\mathfrak{S})$ be the group of all real matrices $\mathfrak{U}$ satisfying $\mathfrak{S}[\mathfrak{U}] = \mathfrak{S}$. By the transformation $\mathfrak{X} \rightarrow \mathfrak{U}\mathfrak{X}$ the space of all non-singular real matrices $\mathfrak{X}^{(m)}$ is mapped onto itself; if $\mathfrak{W}^{(m)}$ is any real symmetric matrix with the same signature as $\mathfrak{S}$, then the $\frac{1}{2}m(m-1)$-dimensional surfaces $\mathfrak{S}[\mathfrak{X}] = \mathfrak{W}$ remain invariant. We consider the m^2 variable elements of $\mathfrak{X}$ as differentiable functions of the $\frac{1}{2}m(m+1)$ independent elements in $\mathfrak{W}$ and of $\frac{1}{2}m(m-1)$ new independent variables $u_1, u_2, \cdots$; let Φ denote the absolute value of the jacobian of this transformation of variables. The formula

$$dv = | \mathfrak{W}\mathfrak{S}^{-1} |^{\frac{1}{2}} \Phi du_1 du_2 \cdots \tag{1}$$

defines a volume element on the surface $\mathfrak{S}[\mathfrak{X}] = \mathfrak{W}$ which is invariant under $\Omega(\mathfrak{S})$; on account of the factor $| \mathfrak{W}\mathfrak{S}^{-1} |^{\frac{1}{2}}$, this volume element does not depend upon $\mathfrak{W}$. If ρ_m denotes the volume of the space of the m-rowed orthogonal matrices $\mathfrak{X}$ obtained for $\mathfrak{S} = \mathfrak{W} = \mathfrak{E}$, then

$$\rho_m = \prod_{k=1}^{m} \frac{\pi^{k/2}}{\Gamma(k/2)}.$$

Let $\mathfrak{G}$ be a given real matrix with m rows and n columns, of rank n, let $\mathfrak{Y}$ be a variable real matrix with m rows and $m-n$ columns, and set $\mathfrak{X} = (\mathfrak{G}, \mathfrak{Y})$. For all $\mathfrak{U}$ in the subgroup $\Omega(\mathfrak{S}, \mathfrak{G})$ of $\Omega(\mathfrak{S})$ defined by the condition $\mathfrak{U}\mathfrak{G} = \mathfrak{G}$, the transformation $\mathfrak{X} \rightarrow \mathfrak{U}\mathfrak{X}$ maps the $\mathfrak{X}$-space onto itself, and the $\frac{1}{2}(m-n)(m-n-1)$-dimensional surfaces $\mathfrak{S}[\mathfrak{X}] = \mathfrak{W}$ remain invariant, $\mathfrak{W}$ denoting any real symmetric matrix of the form

$$\mathfrak{W} = \begin{pmatrix} \mathfrak{T} & \mathfrak{Q} \\ \mathfrak{Q}' & \mathfrak{R} \end{pmatrix}, \qquad \mathfrak{T} = \mathfrak{S}[\mathfrak{G}],$$

with the same signature as $\mathfrak{S}$. The matrices $\mathfrak{Y}$ and $\mathfrak{Q}$, $\mathfrak{R}$ are connected by the equations $\mathfrak{G}'\mathfrak{S}\mathfrak{Y} = \mathfrak{Q}$, $\mathfrak{S}[\mathfrak{Y}] = \mathfrak{R}$. We consider the $m(m-n)$ variable elements of $\mathfrak{Y}$ as differentiable functions of the $\frac{1}{2}(m-n)(m+n+1)$ independent elements in $\mathfrak{Q}$, $\mathfrak{R}$ and of $\frac{1}{2}(m-n)(m-n-1)$ new independent variables u_1, $u_2, \cdots$. If Φ denotes again the absolute value of the jacobian of this transformation, then (1) defines a volume element on the surface $\mathfrak{S}[\mathfrak{X}] = \mathfrak{W}$ which is invariant under $\Omega(\mathfrak{S}, \mathfrak{G})$.

Assume now that $\mathfrak{S}$ and $\mathfrak{G}$ are integral, and denote by $\Gamma(\mathfrak{S})$ and $\Gamma(\mathfrak{S}, \mathfrak{G})$ the subgroups of all integral $\mathfrak{U}$ in $\Omega(\mathfrak{S})$ and $\Omega(\mathfrak{S}, \mathfrak{G})$. Let $\rho(\mathfrak{S})$ be the volume of a fundamental domain on the surface $\mathfrak{S}[\mathfrak{X}] = \mathfrak{W}$ with respect to the discontinuous subgroup $\Gamma(\mathfrak{S})$ of $\Omega(\mathfrak{S})$, computed with the volume element (1), and let $\rho(\mathfrak{S}, \mathfrak{G})$ be the analogous volume for $\Gamma(\mathfrak{S}, \mathfrak{G})$. These volumes are independent of $\mathfrak{W}$.

Two matrices $\mathfrak{G}_1$ and $\mathfrak{G}_2$ are called associate, relative to $\Gamma(\mathfrak{S})$, if there exists at least one element $\mathfrak{U}$ of $\Gamma(\mathfrak{S})$ such that $\mathfrak{G}_2 = \mathfrak{U}\mathfrak{G}_1$. Let $\mathfrak{T}$ be an integral n-rowed symmetric matrix, not necessarily non-singular, and let $\mathfrak{G}$ run over a complete set of non-associate integral solutions of $\mathfrak{S}[\mathfrak{G}] = \mathfrak{T}$, of rank n; then we define

$$(2) \qquad \mu(\mathfrak{S}, \mathfrak{T}) = \sum_{\mathfrak{G}} \rho(\mathfrak{S}, \mathfrak{G})/\rho(\mathfrak{S}),$$

the measure of the representations of $\mathfrak{T}$ by $\mathfrak{S}$. For positive $\mathfrak{S}$ and $\mathfrak{T}$ we have

$$\mu(\mathfrak{S}, \mathfrak{T}) = \frac{\rho_{m-n}}{\rho_m} \,|\, \mathfrak{S} \,|^{n/2} \,|\, \mathfrak{T} \,|^{\frac{1}{2}(n-m+1)} A(\mathfrak{S}, \mathfrak{T}),$$

where $A(\mathfrak{S}, \mathfrak{T})$ is the number of all representations of $\mathfrak{T}$ by $\mathfrak{S}$; consequently, for indefinite quadratic forms, the measure $\mu(\mathfrak{S}, \mathfrak{T})$ is an equivalent of the representation number.

On the other hand, we consider the number $A_q(\mathfrak{S}, \mathfrak{T})$ of modulo q incongruent integral solutions $\mathfrak{G}$ of the congruence $\mathfrak{S}[\mathfrak{G}] \equiv \mathfrak{T} \pmod{q}$, where q is any positive integer. In particular, let $q = p^l (l = 1, 2, \cdots)$ run over all powers of a given prime number p and define

$$\alpha_p(\mathfrak{S}, \mathfrak{T}) = \lim_{l \to \infty} q^{\frac{1}{2}n(n+1)-mn} A_q(\mathfrak{S}, \mathfrak{T}),$$

the p-adic density of the representations of $\mathfrak{T}$ by $\mathfrak{S}$.

Our principal object is the proof of

THEOREM 1. *Let r, $m - r$ be the signature of $\mathfrak{S}$ and let*

$$(3) \qquad n \leqq r, \qquad n \leqq m - r, \qquad 2n + 2 < m;$$

then

$$(4) \qquad \mu(\mathfrak{S}, \mathfrak{T}) = \prod_p \alpha_p(\mathfrak{S}, \mathfrak{T}),$$

where p runs over all primes.

As a consequence of this theorem we shall obtain another theorem concerning the primitive representations of $\mathfrak{T}$ by $\mathfrak{S}$. An integral matrix $\mathfrak{F}^{(m,n)}$ is called primitive if it can be filled up to an m-rowed unimodular matrix; this means that the greatest common divisor of all n-rowed minors of $\mathfrak{F}$ is 1. Let $\mathfrak{F}$ run over a complete system of non-associate primitive solutions of $\mathfrak{S}[\mathfrak{F}] = \mathfrak{T}$, then we define

$$(5) \qquad \nu(\mathfrak{S}, \mathfrak{T}) = \sum_{\mathfrak{F}} \rho(\mathfrak{S}, \mathfrak{F})/\rho(\mathfrak{S}).$$

On the other hand, let $B_q(\mathfrak{S}, \mathfrak{T})$ be the number of modulo q incongruent primitive solutions $\mathfrak{F}$ of the congruence $\mathfrak{S}[\mathfrak{F}] \equiv \mathfrak{T} \pmod{q}$ and define

$$(6) \qquad \beta_p(\mathfrak{S}, \mathfrak{T}) = \lim_{l \to \infty} q^{\frac{1}{2}n(n+1)-mn} B_q(\mathfrak{S}, \mathfrak{T}),$$

where $q = p^l$ runs over all powers of a given prime number p.

THEOREM 2. *If* (3) *is fulfilled, then*

$$\nu(\mathfrak{S}, \mathfrak{T}) = \prod_p \beta_p(\mathfrak{S}, \mathfrak{T}),$$

where p *runs over all primes.*

Theorem 1 is a refinement of the result of a former paper. There I proved: Let $\mathfrak{T}$ be non-singular, let r, $m - r$ and s, $n - s$ be the signatures of $\mathfrak{S}$ and $\mathfrak{T}$, suppose that

$$s \leqq r, \qquad n - s \leqq m - r, \qquad n + 1 < m \tag{7}$$

and let $\mathfrak{S}_1, \cdots, \mathfrak{S}_h$ denote representatives of all classes in the genus of $\mathfrak{S}$; then

$$\sum_{k=1}^{h} \rho(\mathfrak{S}_k)\mu(\mathfrak{S}_k, \mathfrak{T}) \Big/ \sum_{k=1}^{h} \rho(\mathfrak{S}_k) = \prod_p \alpha_p(\mathfrak{S}, \mathfrak{T}). \tag{8}$$

It is obvious that (8) follows from (4) by a summation over the different classes of the genus of $\mathfrak{S}$, provided that the stronger condition (3) instead of (7) is satisfied. Moreover, under the assumption (3), our new result (4) asserts that the quantity $\mu(\mathfrak{S}, \mathfrak{T})$ is a genus invariant; this means that quadratic forms in the same genus have the same representation measures $\mu(\mathfrak{S}, \mathfrak{T})$, whenever $n \leqq r$, $n \leqq m - r$, $2n + 2 < m$. This assertion is trivial if the genus of $\mathfrak{S}$ contains only one class. By a well known theorem of A. Meyer, the class number h is 1, whenever $\mathfrak{S}[\mathfrak{x}]$ is an indefinite quadratic form of more than 2 variables whose determinant does not contain a square factor $\neq 1$. Our result (4) might lead to the hypothesis that each indefinite genus of more than 4 variables contains only one class. However, Witt has discovered an example of two different classes of positive $\mathfrak{S}$ with the same representation numbers $A(\mathfrak{S}, \mathfrak{T})$, for $n = 1, 2$; namely, the two classes of the genus of positive even quadratic forms with 16 variables and determinant 1. Therefore, the formula (4) seems to be rather a weak argument for the truth of the hypothesis.

The proof of Theorem 1 is essentially different from our former proof of (8). We introduce the space H of all positive real symmetric matrices $\mathfrak{H}$ satisfying $\mathfrak{H}\mathfrak{S}^{-1}\mathfrak{H} = \mathfrak{S}$ and the space Z of all complex n-rowed symmetric matrices $\mathfrak{Z} = \mathfrak{X} + i\mathfrak{Y}$ with positive imaginary part $\mathfrak{Y}$. Let $\mathfrak{G}$ run over all integral matrices with m rows and n columns, and define

$$f(\mathfrak{Z}) = f(\mathfrak{Z}, \mathfrak{H}) = \sum_{\mathfrak{G}} e^{-2\pi\sigma(\mathfrak{H}[\mathfrak{G}]\mathfrak{Y} - i\mathfrak{S}[\mathfrak{G}]\mathfrak{X})}.$$

Since $f(\mathfrak{Z}, \mathfrak{H}[\mathfrak{U}]) = f(\mathfrak{Z}, \mathfrak{H})$, for all $\mathfrak{U}$ in $\Gamma(\mathfrak{S})$, the function $f(\mathfrak{Z}, \mathfrak{H})$ is an invariant with respect to the representation $\mathfrak{H} \to \mathfrak{H}[\mathfrak{U}]$ of $\Gamma(\mathfrak{S})$ in H. On the other hand, it is possible to investigate the behavior of $f(\mathfrak{Z})$ for all transformations $\mathfrak{Z} \to (\mathfrak{A}\mathfrak{Z} + \mathfrak{B})(\mathfrak{C}\mathfrak{Z} + \mathfrak{D})^{-1}$ of Z under the modular group of degree n; this is accomplished by the transformation formula in Theorem 3. Furthermore, we generalize the circle method of Hardy and Littlewood; instead of the Farey dissection, we use the properties of the fundamental domain of the modular group of degree n. This leads to the formula of Theorem 4, namely

$$\lim_{\epsilon \to 0} \epsilon^{\frac{1}{2}n(m-n-1)} \sum_{\mathfrak{S}[\mathfrak{G}] = \mathfrak{T}} e^{-\frac{1}{2}\pi\epsilon\sigma(\mathfrak{H}[\mathfrak{G}])} = \frac{\rho_r\, \rho_{m-r}\, \rho_{m-2n-1}}{\rho_{r-n}\, \rho_{m-r-n}\, \rho_{m-n-1}} S^{-n/2} \prod_p \alpha_p(\mathfrak{S}, \mathfrak{T}), \tag{9}$$

where ϵ is a positive number, $\mathfrak{G}$ runs over all integral solutions of $\mathfrak{S}[\mathfrak{G}] = \mathfrak{T}$ and S denotes the absolute value of $|\mathfrak{S}|$. From Theorem 4 we obtain Theorem 1 by integration of (9) over a fundamental domain F of $\Gamma(\mathfrak{S})$ in H; since F is not compact and (9) does not hold uniformly in F, we need a further estimate, given by Theorem 5.

For all positive $\mathfrak{T}$ the formula (4) of Theorem 1 can be expressed in a different way involving modular forms of degree n; this will be indicated in the last chapter.

The conditions (3) are used in several parts of the proof of Theorem 1, and it seems that they cannot be improved very much without changing the whole method. With a considerable number of modifications it is possible to prove (4) in the particular case $n = 1 \leqq r < m = 4$ not contained in (3); consequently, indefinite quaternary quadratic forms in the same genus have the same representation measures $\mu(\mathfrak{S}, t)$, for all numbers t. A sketch of the proof is given in the last chapter.

We add some remarks which are not used in the main part of the paper:

In the notation of E. Cartan, H is a symmetric space with respect to the representation $\mathfrak{H} \to \mathfrak{H}[\mathfrak{U}]$ of $\Omega(\mathfrak{S})$. An invariant line element is defined by the quadratic differential form $ds^2 = \frac{1}{8}\sigma(\mathfrak{H}^{-1}d\mathfrak{H}\mathfrak{H}^{-1}d\mathfrak{H})$; let $v(F)$ denote the volume of the fundamental domain F computed in this metric; then

$$\rho(\mathfrak{S}) = \tfrac{1}{2}\rho_r\,\rho_{m-r}\,S^{-\frac{1}{2}(m+1)}v(F).$$

Consider any subgroup $\Gamma^*(\mathfrak{S})$ of finite index j in $\Gamma(\mathfrak{S})$ with the property that the mappings $\mathfrak{H} \to \mathfrak{H}[\mathfrak{U}]$, for all $\mathfrak{U} \neq \mathfrak{E}$ in $\Gamma^*(\mathfrak{S})$, have no fixed point in H; e.g., this condition is fulfilled for every congruence subgroup of $\Gamma(\mathfrak{S})$ defined by $\mathfrak{U} \equiv \mathfrak{E} \pmod{q}$, where q is an arbitrarily given integer >2. The volume of a fundamental domain F^* of $\Gamma^*(\mathfrak{S})$ on H has the value $v(F^*) = jv(F)$. Identifying all frontier points of F^* which are mapped into each other by transformations of the group, we obtain a Riemannian manifold. Assume that the number $r(m - r)$ of dimensions of H is an even number 2μ. By an application of the formula of Allendoerfer and Weil, the relationship

$$\chi = \pi^{-\mu-\frac{1}{2}}\Gamma(\mu + \tfrac{1}{2})Kv(F^*)$$

is proved, where χ is the characteristic of F^* and the scalar curvature quantity K is a constant, because of the transitivity of $\Omega(\mathfrak{S})$ in H; the value $(-1)^{\mu}K$ is found to be a positive rational number. It follows that

$$\chi = (-1)^{\mu}c_{rm}\,\pi^{-[m^2/4]}S^{\frac{1}{2}(m+1)}j\rho(\mathfrak{S}), \tag{10}$$

where c_{rm} is a positive rational number depending only on r and m.

In a similar way a topological interpretation of $\rho(\mathfrak{S}, \mathfrak{G})$ can be obtained. In particular, let $\mathfrak{G} = \mathfrak{F}$ be primitive and let $\mathfrak{S}[\mathfrak{F}] = \mathfrak{T}$ be non-singular. Completing $\mathfrak{F}$ to a unimodular matrix $\mathfrak{B}$ and setting

$$\mathfrak{B}^{-1}\mathfrak{F} = \mathfrak{F}_1, \qquad \mathfrak{S}[\mathfrak{B}] = \mathfrak{S}_1 = \begin{pmatrix} \mathfrak{T} & \mathfrak{Q} \\ \mathfrak{Q}' & \mathfrak{R} \end{pmatrix}, \qquad \mathfrak{R} - \mathfrak{T}^{-1}[\mathfrak{Q}] = \mathfrak{L},$$

we have $\mathfrak{B}^{-1}\Gamma(\mathfrak{S}, \mathfrak{F})\mathfrak{B} = \Gamma(\mathfrak{S}_1, \mathfrak{F}_1)$, and the elements $\mathfrak{U}$ of $\Gamma(\mathfrak{S}_1, \mathfrak{F}_1)$ are of the form

$$\mathfrak{U} = \begin{pmatrix} \mathfrak{E} & \mathfrak{A} \\ 0 & \mathfrak{B} \end{pmatrix}, \qquad \mathfrak{A} = \mathfrak{T}^{-1}\mathfrak{Q}(\mathfrak{E} - \mathfrak{B}),$$

with unimodular $\mathfrak{B}$ satisfying the conditions

$$\mathfrak{T}^{-1}\mathfrak{Q}(\mathfrak{E} - \mathfrak{B}) \equiv 0 \pmod 1, \qquad \mathfrak{L}[\mathfrak{B}] = \mathfrak{L}.$$

Consequently, the group $\Gamma(\mathfrak{S}, \mathfrak{F})$ is isomorphic to a subgroup of finite index j_1 in $\Gamma(\mathfrak{L})$. The volumes $\rho(\mathfrak{S}, \mathfrak{F})$ and $\rho(\mathfrak{L})$ are related by the formula

$$\rho(\mathfrak{S}, \mathfrak{F}) = T^{n-m} j_1 \rho(\mathfrak{L}), \tag{11}$$

where T denotes the absolute value of $|\mathfrak{T}|$. Let $q, n - q$ be the signature of $\mathfrak{T}$; then $\mathfrak{L}$ has the signature $r - q, m - n - r + q$. If the product $(r - q)(m - n - r + q)$ is even, then (10) and (11) lead to a relationship between $\rho(\mathfrak{S}, \mathfrak{F})$ and the characteristic of a fundamental domain connected with $\Gamma(\mathfrak{S}, \mathfrak{F})$. This shows that Theorem 1 can be interpreted as a formula concerning the characteristics of certain manifolds, whenever the numbers $r(m - r)$ and $(r - q)(m - n - r + q)$ are both even.

The fundamental domain F^* is not compact if $\mathfrak{S}[\mathfrak{x}]$ is a zero form; hence always for $m > 4$. Therefore, the application of the formula of Allendoerfer and Weil to the proof of (10) is not immediate. It is necessary to consider F^* as the limit of a particular sequence of polyhedra, and the passage to the limit requires a detailed study of the points at infinity. This presents no serious difficulty, since the necessary properties of F^* are provided by the theory of reduction, but it is somewhat laborious, and we omit it in the present paper.

2. Algebraic lemmata

For any complex square matrix $\mathfrak{M}$ we denote the absolute value of the determinant $|\mathfrak{M}|$ by abs $\mathfrak{M}$. Let $\mathfrak{S}^{(m)}$ be a non-singular real symmetric matrix with the signature $r, m - r$, and let $\mathfrak{H}$ be a positive real symmetric matrix satisfying $\mathfrak{H}\mathfrak{S}^{-1}\mathfrak{H} = \mathfrak{S}$. Let $\mathfrak{Z} = \mathfrak{X} + i\mathfrak{Y}$ be a complex n-rowed symmetric matrix whose imaginary part $\mathfrak{Y}$ is positive. We introduce a matrix $\mathfrak{B}^{(m,n)} = (\mathfrak{v}_1 \cdots \mathfrak{v}_n)$ with indeterminate elements, set $\mathfrak{v}' = (\mathfrak{v}_1' \cdots \mathfrak{v}_n')$ and define the mn-rowed complex symmetric matrix $\mathfrak{R}$ by the formula

$$\mathfrak{R}[\mathfrak{v}] = \sigma(\mathfrak{H}[\mathfrak{B}]\mathfrak{Y}) - i\mathfrak{S}[\mathfrak{B}]\mathfrak{X}). \tag{12}$$

LEMMA 1.

$$|\mathfrak{R}| = |-i\mathfrak{Z}|^{r}\,|i\bar{\mathfrak{Z}}|^{m-r} \operatorname{abs} \mathfrak{S}^{n}.$$

PROOF. Choose the real matrix $\mathfrak{F}$ such that $\mathfrak{S}[\mathfrak{F}] = \mathfrak{P} = [p_1, \cdots, p_m]$ is a diagonal matrix and $\mathfrak{H}[\mathfrak{F}] = \mathfrak{E}$. Since $\mathfrak{H}\mathfrak{S}^{-1}\mathfrak{H} = \mathfrak{S}$, we obtain $p_k^2 = 1$ ($k = 1, \cdots, m$); hence r of the m diagonal elements p_k are 1 and $m - r$ are -1. Choose the real matrix $\mathfrak{G}$ such that $\mathfrak{X}[\mathfrak{G}] = \mathfrak{Q} = [q_1, \cdots, q_n]$ and $\mathfrak{Y}[\mathfrak{G}] = \mathfrak{E}$.

Define $\mathfrak{V} = \mathfrak{F}\mathfrak{W}\mathfrak{G}'$, $\mathfrak{W} = (w_{kl}) = (\mathfrak{w}_1 \cdots \mathfrak{w}_n)$, $\mathfrak{w}' = (\mathfrak{w}_1' \cdots \mathfrak{w}_n')$, $\mathfrak{F} \times \mathfrak{G} = \mathfrak{M}'$, then $\mathfrak{v} = \mathfrak{M}'\mathfrak{w}$, $|\mathfrak{M}'| = |\mathfrak{F}|^n |\mathfrak{G}|^m$,

$$\mathfrak{H}[\mathfrak{V}]\mathfrak{Y} - i\mathfrak{S}[\mathfrak{V}]\mathfrak{X} = \mathfrak{G}(\mathfrak{W}'\mathfrak{W} - i\mathfrak{W}'\mathfrak{P}\mathfrak{W}\mathfrak{Q})\mathfrak{G}^{-1}$$

$$\mathfrak{N}[\mathfrak{M}'\mathfrak{w}] = \mathfrak{N}[\mathfrak{v}] = \sigma(\mathfrak{W}'\mathfrak{W} - i\mathfrak{W}'\mathfrak{P}\mathfrak{W}\mathfrak{Q}) = \sum_{k=1}^{m}\sum_{l=1}^{n} (1 - ip_k q_l) w_{kl}^2 \tag{13}$$

$$|\mathfrak{N}|\,|\mathfrak{M}'|^2 = \prod_{k,l} (1 - ip_k q_l) = |\mathfrak{E} - i\mathfrak{Q}|^r \, |\mathfrak{E} + i\mathfrak{Q}|^{m-r}$$

$$|\mathfrak{N}|\,|\mathfrak{F}|^{2n} |\mathfrak{G}|^{2m} = |\mathfrak{G}|^{2m} |\mathfrak{Y} - i\mathfrak{X}|^r \, |\mathfrak{Y} + i\mathfrak{X}|^{m-r},$$

moreover $|\mathfrak{S}|\,|\mathfrak{F}|^2 = |\mathfrak{P}| = (-1)^{m-r}$, and the assertion follows.

LEMMA 2. *Let* $-\mathfrak{Z}^{-1} = \mathfrak{X}_1 + i\mathfrak{Y}_1$, *then*

$$\mathfrak{N}^{-1}[\mathfrak{v}] = \sigma(\mathfrak{H}^{-1}[\mathfrak{V}]\mathfrak{Y}_1 - i\mathfrak{S}^{-1}[\mathfrak{V}]\mathfrak{X}_1).$$

PROOF. By (13),

$$\mathfrak{N}^{-1}[\mathfrak{M}'^{-1}\mathfrak{w}] = \sum_{k,l} (1 - ip_k q_l)^{-1} w_{kl}^2 .$$

Since

$$(1 - ip_k q_l)^{-1} = \tfrac{1}{2}\left(\frac{1 + p_k}{1 - iq_l} + \frac{1 - p_k}{1 + iq_l}\right),$$

we obtain

$$\mathfrak{N}^{-1}[\mathfrak{M}'^{-1}\mathfrak{w}] = \tfrac{1}{2}\sigma(\mathfrak{W}'(\mathfrak{E} + \mathfrak{P})\mathfrak{W}(\mathfrak{E} - i\mathfrak{Q})^{-1} + \mathfrak{W}'(\mathfrak{E} - \mathfrak{P})\mathfrak{W}(\mathfrak{E} + i\mathfrak{Q})^{-1}).$$

Performing the substitution $\mathfrak{W} = \mathfrak{F}'\mathfrak{V}\mathfrak{G}$, we have $\mathfrak{w} = \mathfrak{M}\mathfrak{v}$, whence

$$\begin{aligned}\mathfrak{N}^{-1}[\mathfrak{v}] &= \tfrac{1}{2}\sigma(\mathfrak{V}'(\mathfrak{H}^{-1} + \mathfrak{S}^{-1})\mathfrak{V}(\mathfrak{Y} - i\mathfrak{X})^{-1} + \mathfrak{V}'(\mathfrak{H}^{-1} - \mathfrak{S}^{-1})\mathfrak{V}(\mathfrak{Y} + i\mathfrak{X})^{-1}) \\ &= \sigma(\mathfrak{H}^{-1}[\mathfrak{V}]\mathfrak{Y}_1 - i\mathfrak{S}^{-1}[\mathfrak{V}]\mathfrak{X}_1);\end{aligned}$$

q.e.d.

We define $\mathfrak{H}_1 = \tfrac{1}{2}(\mathfrak{S} + \mathfrak{H})$, $\mathfrak{H}_2 = \tfrac{1}{2}(\mathfrak{S} - \mathfrak{H})$.

LEMMA 3. *Let* $\mathfrak{R}_0$, $\mathfrak{R}_1$, $\mathfrak{R}_2$ *be matrices with m rows and h columns and set* $\mathfrak{R} = \mathfrak{R}_0 + \mathfrak{S}^{-1}\mathfrak{H}_1\mathfrak{R}_1 + \mathfrak{S}^{-1}\mathfrak{H}_2\mathfrak{R}_2$, *then*

$$\mathfrak{H}_1[\mathfrak{R}] = \mathfrak{H}_1[\mathfrak{R}_0 + \mathfrak{R}_1], \qquad \mathfrak{H}_2[\mathfrak{R}] = \mathfrak{H}_2[\mathfrak{R}_0 + \mathfrak{R}_2].$$

PROOF. In view of $\mathfrak{H}\mathfrak{S}^{-1}\mathfrak{H} = \mathfrak{S}$, the formulae $\mathfrak{H}_1\mathfrak{S}^{-1}\mathfrak{H}_1 = \mathfrak{H}_1$, $\mathfrak{H}_2\mathfrak{S}^{-1}\mathfrak{H}_2 = \mathfrak{H}_2$, $\mathfrak{H}_1\mathfrak{S}^{-1}\mathfrak{H}_2 = 0$ hold, and the assertion follows from the definition of $\mathfrak{R}$.

LEMMA 4. *Suppose that*

$$\mathfrak{S} = \begin{pmatrix} \mathfrak{S}_1^{(m-r)} & * \\ * & * \end{pmatrix}, \qquad -\mathfrak{S}_1 > 0; \tag{14}$$

then the transformation

$$\begin{pmatrix} \mathfrak{Y} \\ \mathfrak{E} \end{pmatrix} = \mathfrak{X}, \qquad \mathfrak{S}[\mathfrak{X}] = \mathfrak{W} > 0, \qquad 2\mathfrak{W}^{-1}[\mathfrak{X}'\mathfrak{S}] - \mathfrak{S} = \mathfrak{H} \tag{15}$$

maps the $r(m-r)$-dimensional space Y of all real $\mathfrak{Y}^{(m-r,r)}$ with positive $\mathfrak{W}$ onto the space H of all positive symmetric $\mathfrak{H}$ satisfying $\mathfrak{H}\mathfrak{S}^{-1}\mathfrak{H} = \mathfrak{S}$.

PROOF. Let $\mathfrak{X}_0^{(m,r)}$ be real, $\mathfrak{S}[\mathfrak{X}_0] = \mathfrak{W}_0 > 0$ and define $\mathfrak{H} = 2\mathfrak{W}_0^{-1}[\mathfrak{X}_0'\mathfrak{S}] - \mathfrak{S}$, $\mathfrak{H}_1 = \frac{1}{2}(\mathfrak{S} + \mathfrak{H})$, $\mathfrak{H}_2 = \frac{1}{2}(\mathfrak{S} - \mathfrak{H})$; then $\mathfrak{H}_1 = \mathfrak{W}_0^{-1}[\mathfrak{X}_0'\mathfrak{S}]$, whence $\mathfrak{H}_1\mathfrak{S}^{-1}\mathfrak{H}_1 = \mathfrak{H}_1$, $\mathfrak{H}\mathfrak{S}^{-1}\mathfrak{H} = \mathfrak{S}$, $\mathfrak{H}_1\mathfrak{S}^{-1}\mathfrak{H}_2 = 0$, $\mathfrak{H}_2\mathfrak{S}^{-1}\mathfrak{H}_2 = \mathfrak{H}_2$,

$$\mathfrak{S}^{-1}[\mathfrak{H}_1\mathfrak{x}_1 + \mathfrak{H}_2\mathfrak{x}_2] = \mathfrak{H}_1[\mathfrak{x}_1] + \mathfrak{H}_2[\mathfrak{x}_2], \tag{16}$$

with indeterminate columns $\mathfrak{x}_1$, $\mathfrak{x}_2$. Since $\mathfrak{W}_0$ is positive, $\mathfrak{X}_0$ has the rank r and $\mathfrak{H}_1 = \mathfrak{W}_0[\mathfrak{W}_0^{-1}\mathfrak{X}_0'\mathfrak{S}]$ has the signature r, 0; on the other hand, $\mathfrak{S}^{-1} = \mathfrak{S}[\mathfrak{S}^{-1}]$ has the signature r, $m - r$. It follows from (16) that $\mathfrak{H}_2$ is non-positive; hence $\mathfrak{H} = \mathfrak{H}_1 - \mathfrak{H}_2$ is non-negative. But $\mathfrak{H}$ is non-singular, because of $\mathfrak{H}\mathfrak{S}^{-1}\mathfrak{H} = \mathfrak{S}$; therefore, $\mathfrak{H} > 0$. In particular, for $\mathfrak{X}_0 = \mathfrak{X} = \begin{pmatrix}\mathfrak{Y}\\ \mathfrak{E}\end{pmatrix}$, this proves that (15) maps Y into H. Since

$$\tfrac{1}{2}(\mathfrak{H} + \mathfrak{S})[\mathfrak{S}^{-1}] = \mathfrak{H}_1[\mathfrak{S}^{-1}] = \begin{pmatrix}\mathfrak{W}^{-1}[\mathfrak{Y}'] & \mathfrak{Y}\mathfrak{W}^{-1}\\ \mathfrak{W}^{-1}\mathfrak{Y}' & \mathfrak{W}^{-1}\end{pmatrix},$$

different points of Y are mapped into different points of H.

Vice versa, let $\mathfrak{H}$ be an arbitrary point of H, and choose a real $\mathfrak{F}^{(m)}$ such that $\mathfrak{H}[\mathfrak{F}] = \mathfrak{E}$, $\mathfrak{S}[\mathfrak{F}] = [p_1, \cdots, p_m]$; then r of the numbers p_k are 1 and $m - r$ are -1. This implies that $\mathfrak{H}_1 = \frac{1}{2}(\mathfrak{H} + \mathfrak{S})$ is non-negative and of rank r. Using (14) and completing squares, we get

$$\mathfrak{S} = \begin{pmatrix}\mathfrak{S}_1 & 0\\ 0 & \mathfrak{S}_2\end{pmatrix}\begin{bmatrix}\mathfrak{E} & *\\ 0 & \mathfrak{E}\end{bmatrix}, \qquad \mathfrak{S}^{-1} = \begin{pmatrix}\mathfrak{S}_1^{-1} & 0\\ 0 & \mathfrak{S}_2^{-1}\end{pmatrix}\begin{bmatrix}\mathfrak{E} & 0\\ * & \mathfrak{E}\end{bmatrix}, \qquad \mathfrak{S}_2 > 0. \tag{17}$$

If $\mathfrak{x}$ is a real column whose first $m - r$ elements are 0, then $\mathfrak{S}^{-1}[\mathfrak{x}] \geqq 0$, by (17); therefore $\mathfrak{H}_1[\mathfrak{S}^{-1}\mathfrak{x}] = \frac{1}{2}\mathfrak{H}[\mathfrak{S}^{-1}\mathfrak{x}] + \frac{1}{2}\mathfrak{S}^{-1}[\mathfrak{x}] > 0$, except for $\mathfrak{x} = 0$. This proves that the matrix $\mathfrak{H}_0^{(r)}$ in

$$\mathfrak{H}_1[\mathfrak{S}^{-1}] = \begin{pmatrix}* & *\\ * & \mathfrak{H}_0\end{pmatrix}$$

is non-singular. Set $\mathfrak{H}_0^{-1} = \mathfrak{W}$, then $\mathfrak{W} > 0$ and

$$\mathfrak{H}_1[\mathfrak{S}^{-1}] = \begin{pmatrix}0 & 0\\ 0 & \mathfrak{H}_0\end{pmatrix}\begin{bmatrix}\mathfrak{E} & 0\\ * & \mathfrak{E}\end{bmatrix} = \mathfrak{W}^{-1}[*, \mathfrak{E}] = \mathfrak{W}^{-1}[\mathfrak{X}'], \qquad \mathfrak{X} = \begin{pmatrix}\mathfrak{Y}\\ \mathfrak{E}\end{pmatrix},$$

with real $\mathfrak{Y}^{(m-r,r)}$. Furthermore, $\mathfrak{H}_1\mathfrak{S}^{-1}\mathfrak{H}_1 = \mathfrak{H}_1$ implies

$$(\mathfrak{S}[\mathfrak{X}] - \mathfrak{W})[\mathfrak{W}^{-1}\mathfrak{X}'] = 0,$$

whence $\mathfrak{S}[\mathfrak{X}] = \mathfrak{W}$. Because of $\mathfrak{H} = 2\mathfrak{H}_1 - \mathfrak{S}$, the equations (15) are fulfilled. This proves that every point of H is represented by (15).

3. Modular substitutions

All statements in this chapter are known, except Lemma 8; proofs are contained in my papers: *Ueber die analytische Theorie der quadratischen Formen,*

Ann. of Math. (2) 36, pp. 527–606 (1935); *Einführung in die Theorie der Modulfunktionen n-ten Grades*, Math. Ann. 116, pp. 617–657 (1939).

Let $\mathfrak{R}$ be a rational n-rowed symmetric matrix. There exists an n-rowed matrix $\mathfrak{C}$ such that the matrix $(\mathfrak{C}, \mathfrak{D})$, with $\mathfrak{D} = \mathfrak{C}\mathfrak{R}$, is primitive; then $\mathfrak{R} = \mathfrak{C}^{-1}\mathfrak{D}$, and $\mathfrak{C}$, $\mathfrak{D}$ are integral, $|\,\mathfrak{C}\,| \neq 0$, $\mathfrak{C}\mathfrak{D}' = \mathfrak{D}\mathfrak{C}'$. The matrix $\mathfrak{C}$ is called denominator of $\mathfrak{R}$, it is determined up to an arbitrary unimodular factor $\mathfrak{U}^{(n)}$ on the left side; we may choose $\mathfrak{U}$ such that $|\,\mathfrak{U}\mathfrak{C}\,| > 0$. On the other hand, if two n-rowed matrices $\mathfrak{C}$, $\mathfrak{D}$ are given with $|\,\mathfrak{C}\,| \neq 0$, $\mathfrak{C}\mathfrak{D}' = \mathfrak{D}\mathfrak{C}'$ and primitive $(\mathfrak{C}, \mathfrak{D})$, then $\mathfrak{C}$ is denominator of the symmetric matrix $\mathfrak{R} = \mathfrak{C}^{-1}\mathfrak{D}$. We define abs $\mathfrak{C} = \lceil \mathfrak{R} \rceil$.

More generally, consider any two n-rowed matrices $\mathfrak{C}$, $\mathfrak{D}$ with $\mathfrak{C}\mathfrak{D}' = \mathfrak{D}\mathfrak{C}'$ and primitive $(\mathfrak{C}, \mathfrak{D})$; they constitute a coprime symmetric n-pair. We say that two such pairs $\mathfrak{C}$, $\mathfrak{D}$ and $\mathfrak{C}_1$, $\mathfrak{D}_1$ belong to the same class, whenever $\mathfrak{C}\mathfrak{D}_1' = \mathfrak{D}\mathfrak{C}_1'$; this occurs, if and only if $(\mathfrak{C}_1, \mathfrak{D}_1) = \mathfrak{U}(\mathfrak{C}, \mathfrak{D})$ with suitably chosen unimodular $\mathfrak{U}$. Plainly, the expression abs $(\mathfrak{C}\mathfrak{Z} + \mathfrak{D})$ depends only on the class of the pair $\mathfrak{C}$, $\mathfrak{D}$, for any given complex matrix $\mathfrak{Z}^{(n)}$.

In each class of coprime symmetric n-pairs $\mathfrak{C}$, $\mathfrak{D}$, the rank of $\mathfrak{C}$ is fixed; if this rank is h, then the class is called an h-class. There exists only one 0-class, the class of 0, $\mathfrak{E}$. On the other hand, for every n-class, the matrix $\mathfrak{C}^{-1}\mathfrak{D} = \mathfrak{R}$ is fixed; *vice versa*, each rational symmetric $\mathfrak{R}^{(n)}$ determines a single n-class.

We say that two matrices $\mathfrak{F}$ and $\mathfrak{F}_1$ are left-equivalent, whenever $\mathfrak{F}_1 = \mathfrak{F}\mathfrak{V}$ with unimodular $\mathfrak{V}$. Once for all we choose a complete set F_h of non-equivalent primitive matrices $\mathfrak{F}^{(n,h)}$ $(h = 1, \cdots, n)$; in particular, $F_n = \mathfrak{E}$. For each $\mathfrak{F}^{(n,h)}$ with given $h < n$, we determine a fixed complement, i.e., a primitive matrix $\mathfrak{F}^{*(n,n-h)}$ such that $(\mathfrak{F}, \mathfrak{F}^*) = \mathfrak{U}$ is unimodular and $|\,\mathfrak{U}\,| = 1$. Let U_h denote the set of these $\mathfrak{U}$, and define $U_n = \mathfrak{E}$. Furthermore, we choose a fixed denominator $\mathfrak{C}_0$ with $|\,\mathfrak{C}_0\,| > 0$ for each rational h-rowed symmetric matrix $\mathfrak{R}_0$ and put $\mathfrak{D}_0 = \mathfrak{C}_0\mathfrak{R}_0$; let C_h be the set of all pairs $\mathfrak{C}_0$, $\mathfrak{D}_0$.

LEMMA 5. *Let $\mathfrak{U}$ run over U_h and $\mathfrak{C}_0$, $\mathfrak{D}_0$ over C_h; then the pairs of n-rowed matrices*

$$\mathfrak{C} = \begin{pmatrix} \mathfrak{C}_0 & 0 \\ 0 & 0 \end{pmatrix} \mathfrak{U}', \qquad \mathfrak{D} = \begin{pmatrix} \mathfrak{D}_0 & 0 \\ 0 & \mathfrak{E} \end{pmatrix} \mathfrak{U}^{-1}$$

represent in one and only one way all h-classes of coprime symmetric n-pairs. If $\mathfrak{U} = (\mathfrak{F}, \mathfrak{F}^*)$ and $\mathfrak{C}_0^{-1}\mathfrak{D}_0 = \mathfrak{R}_0$, then

$$|\,\mathfrak{C}\mathfrak{Z} + \mathfrak{D}\,| = \lceil \mathfrak{R}_0 \rceil \, |\,\mathfrak{Z}[\mathfrak{F}] + \mathfrak{R}_0\,|,$$

for any complex symmetric $\mathfrak{Z}^{(n)}$.

Let $\mathfrak{Z}$ be a variable complex n-rowed symmetric matrix with positive imaginary part. A modular substitution $\mathfrak{Z} \to (\mathfrak{A}_1\mathfrak{Z} + \mathfrak{B}_1)(\mathfrak{C}_1\mathfrak{Z} + \mathfrak{D}_1)^{-1}$ of degree n is defined by the conditions

$$\mathfrak{A}_1\mathfrak{B}_1' = \mathfrak{B}_1\mathfrak{A}_1', \qquad \mathfrak{C}_1\mathfrak{D}_1' = \mathfrak{D}_1\mathfrak{C}_1', \qquad \mathfrak{A}_1\mathfrak{D}_1' - \mathfrak{B}_1\mathfrak{C}_1' = \mathfrak{E}, \tag{18}$$

with integral n-rowed matrices $\mathfrak{A}_1$, $\mathfrak{B}_1$, $\mathfrak{C}_1$, $\mathfrak{D}_1$. Because of the second and third of these conditions, the matrices $\mathfrak{C}_1$, $\mathfrak{D}_1$ form a coprime symmetric n-pair.

Vice versa, for any coprime symmetric n-pair $\mathfrak{C}_1$, $\mathfrak{D}_1$, there exist two integral matrices $\mathfrak{A}_1$, $\mathfrak{B}_1$ satisfying the first and third of the conditions (18); this means that $\mathfrak{C}_1$ and $\mathfrak{D}_1$ are matrix coefficients in a suitably chosen modular substitution of degree n.

The modular substitutions constitute a group M; the inverse of $\mathfrak{Z} \to (\mathfrak{A}_1\mathfrak{Z} + \mathfrak{B}_1)(\mathfrak{C}_1\mathfrak{Z} + \mathfrak{D}_1)^{-1}$ is $\mathfrak{Z} \to (\mathfrak{D}_1'\mathfrak{Z} - \mathfrak{B}_1')(\mathfrak{A}_1' - \mathfrak{C}_1'\mathfrak{Z})^{-1}$, whence $\mathfrak{A}_1'\mathfrak{C}_1 = \mathfrak{C}_1'\mathfrak{A}_1$, $\mathfrak{B}_1'\mathfrak{D}_1 = \mathfrak{D}_1'\mathfrak{B}_1$, $\mathfrak{A}_1'\mathfrak{D}_1 - \mathfrak{C}_1'\mathfrak{B}_1 = \mathfrak{E}$. The integral modular substitutions are defined by the condition $\mathfrak{C}_1 = 0$; they have the form $\mathfrak{Z} \to \mathfrak{Z}[\mathfrak{B}] + \mathfrak{T}$ with arbitrary unimodular $\mathfrak{B}^{(n)}$ and integral symmetric $\mathfrak{T}^{(n)}$, and they constitute a subgroup M_0 of M. Two modular substitutions with the matrix coefficients $\mathfrak{A}_1$, $\mathfrak{B}_1$, $\mathfrak{C}_1$, $\mathfrak{D}_1$ and $\mathfrak{A}_2$, $\mathfrak{B}_2$, $\mathfrak{C}_2$, $\mathfrak{D}_2$ lie then and only then in the same right coset of M_0 relative to M, if the two pairs $\mathfrak{C}_1$, $\mathfrak{D}_1$ and $\mathfrak{C}_2$, $\mathfrak{D}_2$ belong to the same class. It follows that each class of coprime symmetric n-pairs $\mathfrak{C}$, $\mathfrak{D}$ determines one and only one right coset, and we have

LEMMA 6. *If $\mathfrak{A}_1$, $\mathfrak{B}_1$, $\mathfrak{C}_1$, $\mathfrak{D}_1$ are the coefficients of a given modular substitution of degree n, then the transformation*

$$(\mathfrak{C}, \mathfrak{D}) \to (\mathfrak{C}, \mathfrak{D})\begin{pmatrix}\mathfrak{A}_1 & \mathfrak{B}_1\\ \mathfrak{C}_1 & \mathfrak{D}_1\end{pmatrix}$$

maps the set of all classes of coprime symmetric n-pairs $\mathfrak{C}$, $\mathfrak{D}$ onto itself.

For each pair $\mathfrak{C}_0$, $\mathfrak{D}_0$ in the set C_h, we choose two fixed h-rowed matrices $\mathfrak{A}_0$, $\mathfrak{B}_0$ such that $\mathfrak{A}_0$, $\mathfrak{B}_0$, $\mathfrak{C}_0$, $\mathfrak{D}_0$ are the coefficients of a modular substitution of degree h. Let $\mathfrak{U}$ lie in the set U_h; then the matrices

$$\mathfrak{A} = \begin{pmatrix}\mathfrak{A}_0 & 0\\ 0 & \mathfrak{E}\end{pmatrix}\mathfrak{U}', \qquad \mathfrak{B} = \begin{pmatrix}\mathfrak{B}_0 & 0\\ 0 & 0\end{pmatrix}\mathfrak{U}^{-1}, \qquad \mathfrak{C} = \begin{pmatrix}\mathfrak{C}_0 & 0\\ 0 & 0\end{pmatrix}\mathfrak{U}', \qquad \mathfrak{D} = \begin{pmatrix}\mathfrak{D}_0 & 0\\ 0 & \mathfrak{E}\end{pmatrix}\mathfrak{U}^{-1}$$

are the coefficients of a modular substitution of degree n. This modular substitution shall be called reduced of type h; it is uniquely determined by $\mathfrak{F}$ and $\mathfrak{R}_0 = \mathfrak{C}_0^{-1}\mathfrak{D}_0$. The identical substitution is called reduced of type 0. As a consequence of Lemma 5 we have

LEMMA 7. *Every modular substitution $\mathfrak{Z} \to \mathfrak{Z}^{**}$ is the product of an integral modular substitution $\mathfrak{Z}^{**} = \mathfrak{Z}^*[\mathfrak{B}] + \mathfrak{T}$ and a reduced modular substitution $\mathfrak{Z}^* = (\mathfrak{A}\mathfrak{Z} + \mathfrak{B})(\mathfrak{C}\mathfrak{Z} + \mathfrak{D})^{-1}$, in one and only one way.*

We put

$$\mathfrak{Z}^* = \begin{pmatrix}\mathfrak{Z}_0 & \mathfrak{Z}_{01}\\ \mathfrak{Z}_{01}' & \mathfrak{Z}_1\end{pmatrix},$$

with h-rowed $\mathfrak{Z}_0$.

LEMMA 8. *Let the modular substitution $\mathfrak{Z}^* = (\mathfrak{A}\mathfrak{Z} + \mathfrak{B})(\mathfrak{C}\mathfrak{Z} + \mathfrak{D})^{-1}$ be reduced of type h; then*

$$\mathfrak{Z}[\mathfrak{U}] = \begin{pmatrix}-\mathfrak{C}_0^{-1}\mathfrak{D}_0 & 0\\ 0 & \mathfrak{Z}_1\end{pmatrix} - (\mathfrak{Z}_0 - \mathfrak{A}_0\mathfrak{C}_0^{-1})^{-1}[\mathfrak{C}_0'^{-1}, \mathfrak{Z}_{01}].$$

Proof. We have

$$(19)\quad (\mathfrak{A} - \mathfrak{Z}^*\mathfrak{C})\mathfrak{U}'^{-1} = \begin{pmatrix} \mathfrak{A}_0 - \mathfrak{Z}_0\mathfrak{C}_0 & 0 \\ -\mathfrak{Z}_{01}'\mathfrak{C}_0 & \mathfrak{E} \end{pmatrix}, \quad (\mathfrak{Z}^*\mathfrak{D} - \mathfrak{B})\mathfrak{U} = \begin{pmatrix} \mathfrak{Z}_0\mathfrak{D}_0 - \mathfrak{B}_0 & \mathfrak{Z}_{01} \\ \mathfrak{Z}_{01}'\mathfrak{D}_0 & \mathfrak{Z}_1 \end{pmatrix}$$

$$\mathfrak{U}'(\mathfrak{A} - \mathfrak{Z}^*\mathfrak{C})^{-1} = \begin{pmatrix} \mathfrak{E} & 0 \\ \mathfrak{Z}_{01}'\mathfrak{C}_0 & \mathfrak{E} \end{pmatrix} \begin{pmatrix} (\mathfrak{A}_0 - \mathfrak{Z}_0\mathfrak{C}_0)^{-1} & 0 \\ 0 & \mathfrak{E} \end{pmatrix}$$

$$(\mathfrak{Z}^*\mathfrak{D} - \mathfrak{B})\mathfrak{U} - (\mathfrak{A} - \mathfrak{Z}^*\mathfrak{C})\mathfrak{U}'^{-1} \begin{pmatrix} -\mathfrak{C}_0^{-1}\mathfrak{D}_0 & 0 \\ 0 & \mathfrak{Z}_1 \end{pmatrix} = \begin{pmatrix} \mathfrak{A}_0\mathfrak{C}_0^{-1}\mathfrak{D}_0 - \mathfrak{B}_0 & \mathfrak{Z}_{01} \\ 0 & 0 \end{pmatrix}.$$

Since $\mathfrak{A}_0\mathfrak{C}_0^{-1}\mathfrak{D}_0 - \mathfrak{B}_0 = (\mathfrak{A}_0\mathfrak{D}_0' - \mathfrak{B}_0\mathfrak{C}_0')\mathfrak{C}_0'^{-1} = \mathfrak{C}_0'^{-1}$ and $\mathfrak{Z} = (\mathfrak{A} - \mathfrak{Z}^*\mathfrak{C})^{-1}(\mathfrak{Z}^*\mathfrak{D} - \mathfrak{B})$, we obtain

$$\mathfrak{Z}[\mathfrak{U}] - \begin{pmatrix} -\mathfrak{C}_0^{-1}\mathfrak{D}_0 & 0 \\ 0 & \mathfrak{Z}_1 \end{pmatrix} = \begin{pmatrix} \mathfrak{E} & 0 \\ \mathfrak{Z}_{01}'\mathfrak{C}_0 & \mathfrak{E} \end{pmatrix} \begin{pmatrix} (\mathfrak{A}_0 - \mathfrak{Z}_0\mathfrak{C}_0)^{-1} & 0 \\ 0 & \mathfrak{E} \end{pmatrix} \begin{pmatrix} \mathfrak{C}_0'^{-1} & \mathfrak{Z}_{01} \\ 0 & 0 \end{pmatrix}$$

$$= (\mathfrak{A}_0\mathfrak{C}_0^{-1} - \mathfrak{Z}_0)^{-1}[\mathfrak{C}_0'^{-1}, \mathfrak{Z}_{01}];$$

q.e.d.

4. Arithmetic lemmata

Let $c_1, \cdots, c_n$ be positive integers, $c_k \mid c_{k+1}$ $(k = 1, \cdots, n-1)$ and define the n-rowed diagonal matrix $\mathfrak{K} = [c_1, \cdots, c_n]$. Let U be the group of all n-rowed unimodular matrices $\mathfrak{U}$, and let K be the subgroup of all $\mathfrak{U}$ with integral $\mathfrak{K}\mathfrak{U}\mathfrak{K}^{-1}$.

Lemma 9. *The index of K in U fulfills the inequality*

$$[U:K] \leqq \prod_{p \mid c_n} (1 - p^{-1})^{1-n} \prod_{k=1}^{n} c_k^{2k-n-1},$$

where p runs over all different prime factors of c_n.

Proof. Let q be any positive multiple of c_n, and let Q be the invariant subgroup of U consisting of all $U \equiv \mathfrak{E} \pmod{q}$. Since Q is also a subgroup of K, we have $[U:K] = [U/Q:K/Q]$. The factor group U/Q is isomorphic to the group of all integral n-rowed matrices $\mathfrak{B}$ modulo q with $|\mathfrak{B}| \equiv \pm 1 \pmod{q}$. If q_1, q_2 are coprime positive integers, then the ring of residue classes modulo q_1q_2 is the direct sum of the rings of residue classes modulo q_1 and q_2. Consequently, it suffices to prove the inequality

$$[U^*:K^*] \leqq (1 - p^{-1})^{1-n} \prod_{k=1}^{n} c_k^{2k-n-1},$$

where q is a power of the prime number p and a multiple of c_n, the group U^* consists of all integral n-rowed matrices $\mathfrak{B}$ modulo q with $|\mathfrak{B}| \equiv \pm 1 \pmod{q}$ and K^* is the subgroup of all $\mathfrak{B}$ with integral $\mathfrak{K}\mathfrak{B}\mathfrak{K}^{-1}$.

Furthermore, let V_n be the group of integral $\mathfrak{B}^{(n)}$ modulo q with $(|\mathfrak{B}|, q) = 1$, and let K_n be the subgroup of all $\mathfrak{B}$ in V_n with integral $\mathfrak{K}\mathfrak{B}\mathfrak{K}^{-1}$; plainly, $[V_n:U^*] =$

$[K_n : K^*]$, whence $[V_n : K_n] = [U^* : K^*]$. If $[V_n]$ and $[K_n]$ denote the orders of V_n and K_n, then it suffices to prove that

$$[K_n] \geqq [V_n](1 - p^{-1})^{n-1} \prod_{k=1}^{n} c_k^{n-2k+1}. \tag{20}$$

It is well known that

$$[V_n] = q^{n^2} \prod_{k=1}^{n} (1 - p^{-k}). \tag{21}$$

In the special case $c_1 = c_n$, we have $\mathfrak{R} = c_1\mathfrak{E}$; then $K_n = V_n$, and (20) is true, because of $\sum_{k=1}^{n} (n + 1 - 2k) = 0$; this holds in particular for $n = 1$. We apply induction with respect to n, and we may suppose that $c_1 < c_n$. Define h by the condition $c_h < c_{h+1} = c_n$, then $1 \leqq h \leqq n - 1$. We put

$$\mathfrak{B} = (v_{kl}) = \begin{pmatrix} \mathfrak{B}_1 & \mathfrak{B}_2 \\ \mathfrak{B}_3 & \mathfrak{B}_4 \end{pmatrix},$$

with h-rowed $\mathfrak{B}_1$. The matrices $\mathfrak{B}$ and $\mathfrak{R}\mathfrak{B}\mathfrak{R}^{-1}$ are both integral, if and only if v_{kl} and $c_k v_{kl} c_l^{-1}$ are integers for $k, l = 1, \cdots, n$; then $\mathfrak{B}_3$ and $\mathfrak{B}_4$ are arbitrary integral matrices, whereas $\mathfrak{B}_1$ and $\mathfrak{B}_2$ are integral matrices subjected to the conditions $c_k^{-1} c_l \mid v_{kl}$ $(k \leqq h, k < l)$. Since $p \mid c_k^{-1} c_l$ for $k \leqq h < l$, we infer that $\mathfrak{B}_2 \equiv 0 \pmod{p}$, whence $|\mathfrak{B}| \equiv |\mathfrak{B}_1||\mathfrak{B}_4| \pmod{p}$. Consequently, we get the elements $\mathfrak{B}$ of K_n in the following way: $\mathfrak{B}_4$ is any element of V_{n-h}; $\mathfrak{B}_3$ is an arbitrary integral matrix modulo q; $\mathfrak{B}_2$ is any matrix modulo q satisfying the conditions $c_k^{-1} c_l \mid v_{kl}$ $(k \leqq h < l)$; $\mathfrak{B}_1$ is any element of K_h. It follows that

$$[K_n] = a q^{h(n-h)} [V_{n-h}][K_h],$$

where a is the number of matrices $\mathfrak{B}_2$, namely

$$a = q^{h(n-h)} \prod_{k \leqq h < l} (c_k c_l^{-1}).$$

Applying (20) with h instead of n and (21) with $h, n - h$ instead of n, we obtain

$$[K_n] \geqq q^{n^2}(1 - p^{-1})^{h-1} \prod_{k=1}^{h} c_k^{h-2k+1} \prod_{k \leqq h < l} (c_k c_l^{-1}) \prod_{k=1}^{h} (1 - p^{-k}) \prod_{k=1}^{n-h} (1 - p^{-k}).$$

Since

$$q^{n^2} \prod_{k=1}^{h} (1 - p^{-k}) > [V_n], \qquad \prod_{k=1}^{n-h} (1 - p^{-k}) \geqq (1 - p^{-1})^{n-h},$$

$$\prod_{k=1}^{h} c_k^{h-2k+1} \prod_{k \leqq h < l} (c_k c_l^{-1}) = c_n^{-h(n-h)} \prod_{k=1}^{h} c_k^{n-2k+1} = \prod_{k=1}^{n} c_k^{n-2k+1},$$

the assertion (20) follows.

LEMMA 10. *Let $A(c_1, \cdots, c_n)$ denote the number of modulo 1 incongruent rational n-rowed symmetric matrices whose denominators have the given elementary divisors $c_1, \cdots, c_n$; then*

$$A(c_1, \cdots . c_n) \leqq \prod_{p \mid c_n} (1 - p^{-1})^{1-n} \prod_{k=1}^{n} c_k^k .$$

PROOF. Let $\mathfrak{C}$ be any integral n-rowed matrix with the elementary divisors $c_1, \cdots, c_n$. Choose two unimodular matrices $\mathfrak{U}_0$ and $\mathfrak{U}$ such that $\mathfrak{U}_0^{-1}\mathfrak{C}\mathfrak{U}^{-1} = \mathfrak{K} = [c_1, \cdots, c_n]$. Let $A(\mathfrak{C})$ denote the number of modulo 1 incongruent symmetric $\mathfrak{R}$ with integral $\mathfrak{C}\mathfrak{R}$, and set $\mathfrak{R}[\mathfrak{U}'] = \mathfrak{R}_1 = (r_{kl})$; then $\mathfrak{C}\mathfrak{R}\mathfrak{U}' = \mathfrak{U}_0\mathfrak{K}\mathfrak{R}_1$, whence $A(\mathfrak{C}) = A(\mathfrak{K})$. The matrix $\mathfrak{K}\mathfrak{R}_1$ is integral, if and only if $c_k r_{kl}$ is an integer for $k, l = 1, \cdots, n$. Since $r_{kl} = r_{lk}$ and $c_1 \mid c_2 \mid \cdots \mid c_n$, we infer that

$$A(\mathfrak{K}) = \prod_{k=1}^{n} c_k^{n-k+1}. \tag{22}$$

Plainly, the number of modulo 1 incongruent symmetric $\mathfrak{R}$ with the same denominator $\mathfrak{C}$ is at most $A(\mathfrak{C})$. On the other hand, the matrices $\mathfrak{C} = \mathfrak{U}_0\mathfrak{K}\mathfrak{U}$ and $\mathfrak{C}^* = \mathfrak{U}_0^*\mathfrak{K}\mathfrak{U}^*$, with unimodular $\mathfrak{U}_0^*$ and $\mathfrak{U}^*$, are then and only then denominators of the same $\mathfrak{R}$, if $\mathfrak{C}^*\mathfrak{C}^{-1}$ is unimodular; this means that $\mathfrak{K}\mathfrak{U}^*\mathfrak{U}^{-1}\mathfrak{K}^{-1}$ is integral, $\mathfrak{U}^*\mathfrak{U}^{-1}$ belongs to the subgroup K of the unimodular group U, and $\mathfrak{U}$, $\mathfrak{U}^*$ lie in the same right coset of K in U. Consequently,

$$A(c_1, \cdots, c_n) \leqq [U:K]A(\mathfrak{K}),$$

and the assertion follows from (22) and Lemma 9.

LEMMA 11. *Let $\mathfrak{R}^{(n)}$ run over a complete system of modulo 1 incongruent rational symmetric matrices; then the Dirichlet series*

$$\psi(s) = \sum_{\mathfrak{R}(\text{mod } 1)} \overline{|\mathfrak{R}|}^{-n-s}$$

converges for $s > 1$. If $u > 0$ and $s > 1$, then

$$u^{-s} \sum_{\overline{|\mathfrak{R}|} < u} \overline{|\mathfrak{R}|}^{-n} + \sum_{\overline{|\mathfrak{R}|} \geqq u} \overline{|\mathfrak{R}|}^{-n-s} < a\left(2 + \frac{1}{s-1}\right)u^{1-s},$$

where a depends only on n.

PROOF. If $\lambda(s)$, $\mu(s)$ are two Dirichlet series with non-negative coefficients l_k, m_k satisfying $l_k \leqq m_k$ $(k = 1, 2, \cdots)$, then we write $\lambda(s) \prec \mu(s)$.

In view of the definition of $A(c_1, \cdots, c_n)$, we have

$$\psi(s) = \sum_{c_1 \mid c_2 \mid \cdots \mid c_n} A(c_1, \cdots, c_n)(c_1 \cdots c_n)^{-n-s},$$

where $c_1, \cdots, c_n$ run over all systems of positive integers fulfilling the condition $c_1 \mid c_2 \mid \cdots \mid c_n$. Using Lemma 10 and letting $c_1, \cdots, c_n$ run independently over all positive integers, we get

$$\psi(s) \prec \sum_{c_1, \cdots, c_n} \prod_{p \mid c_n} (1 - p^{-1})^{1-n} \prod_{k=1}^{n} c_k^{k-n-s}$$

$$= \prod_{p} \left(1 + (1 - p^{-1})^{1-n} \sum_{l=1}^{\infty} p^{-ls}\right) \prod_{k=1}^{n-1} \zeta(s + n - k).$$

Put $2^n + n - 3 = \nu$, $\zeta^\nu(s+1) = \gamma(s)$, $(1 - p^{-1})^{1-n} - 1 = p^{-1}b_p$; then $0 \leqq b_p \leqq 2^n - 2 = \nu - n + 1$, for all $p \geqq 2$, and

$$1 + (1 - p^{-1})^{1-n} \sum_{l=1}^{\infty} p^{-ls} = (1 + b_p p^{-1-s})(1 - p^{-s})^{-1}$$
$$< (1 - p^{-1-s})^{n-\nu-1}(1 - p^{-s})^{-1},$$

whence

$$\psi(s) < \gamma(s)\zeta(s). \tag{23}$$

This proves the first assertion of the lemma.

Let a_k, d_k be the coefficients of the Dirichlet series $\psi(s)$, $\gamma(s)$ and set $\sum_{l=1}^{k} a_l = \sigma_k$, $\gamma(1) = \zeta^\nu(2) = a$. By (23),

$$\sigma_k \leqq \sum_{l=1}^{k} d_l \left[\frac{k}{l}\right] \leqq k \sum_{l=1}^{k} d_l l^{-1} < k \sum_{l=1}^{\infty} d_l l^{-1} = ak \qquad (k = 1, 2, \cdots);$$

hence

$$\sum_{|\mathfrak{R}| < u} |\mathfrak{R}|^{-n} = \sum_{l<u} a_l < au, \tag{24}$$

for all positive u. Moreover, for $s > 1$,

$$\sum_{|\mathfrak{R}| \geqq u} |\mathfrak{R}|^{-n-s} = \sum_{k \geqq u} a_k k^{-s} = \sum_{k \geqq u} (\sigma_k - \sigma_{k-1})k^{-s} \leqq \sum_{k \geqq u} \sigma_k(k^{-s} - (k+1)^{-s})$$

$$= s \sum_{k \geqq u} \sigma_k \int_k^{k+1} x^{-s-1}\,dx < as \sum_{k \geqq u} \int_k^{k+1} x^{-s}\,dx \leqq as \int_u^{\infty} x^{-s}\,dx = \frac{as}{s-1} u^{1-s}. \tag{25}$$

The second assertion of the lemma follows from (24) and (25).

Let $\mathfrak{P}^{(g)}$ be a positive real symmetric matrix, reduced in the sense of Minkowski; let $p_1, \cdots, p_g$ be the diagonal elements of $\mathfrak{P}$; then $p_k \leqq p_{k+1}$ ($k = 1, \cdots, g - 1$) and

$$\prod_{k=1}^{g} p_k < b_0 |\mathfrak{P}|, \tag{26}$$

where b_0 depends only on g, by a well known theorem of Hermite and Minkowski. We define the diagonal matrix $\mathfrak{P}_0 = [p_1, \cdots, p_g]$.

LEMMA 12. *There exists a positive number $b = b(g)$ depending only on g such that $\mathfrak{P} > b\mathfrak{P}_0$.*

PROOF. Set $\mathfrak{P}_0^{\frac{1}{2}} = [p_1^{\frac{1}{2}}, \cdots, p_g^{\frac{1}{2}}]$; then all elements of $\mathfrak{P}[\mathfrak{P}_0^{-\frac{1}{2}}]$ have the absolute value $\leqq 1$, and $|\mathfrak{P}[\mathfrak{P}_0^{-\frac{1}{2}}]| > b_0^{-1}$, by (26). Let $\lambda_1, \cdots, \lambda_g$ be the roots of the equation $|\lambda\mathfrak{P}_0 - \mathfrak{P}| = 0$, with $\lambda_1 \leqq \lambda_k$ ($k = 2, \cdots, g$); since $|\lambda\mathfrak{P}_0 - \mathfrak{P}| = |\mathfrak{P}_0|\,|\lambda\mathfrak{E} - \mathfrak{P}[\mathfrak{P}_0^{-\frac{1}{2}}]|$, these roots are bounded positive numbers, the bound depending only on g, and their product is $> b_0^{-1}$. It follows that λ_1 has a positive lower bound. On the other hand, the smallest root λ_1 is the least upper bound of all real λ satisfying $\mathfrak{P} > \lambda\mathfrak{P}_0$. This proves the assertion.

Let also $\mathfrak{Q}^{(h)}$ be positive real symmetric and reduced, with the diagonal elements $q_1, \cdots, q_h$, and set $\mathfrak{Q}_0 = [q_1, \cdots, q_h]$.

LEMMA 13. *Let $\mathfrak{W}^{(g,h)}$ be a real matrix, then*

$$\sigma(\mathfrak{P}[\mathfrak{W}]\mathfrak{Q}) \geqq b_1\sigma(\mathfrak{P}_0[\mathfrak{W}]\mathfrak{Q}_0) \geqq b_1q_1\sigma(\mathfrak{P}_0[\mathfrak{W}]) \geqq b_1p_1q_1\sigma([\mathfrak{W}]),$$

where $b_1 = b(g)b(h)$.

PROOF. Choose a real matrix $\mathfrak{L}$ such that $\mathfrak{Q} = \mathfrak{L}\mathfrak{L}'$. By Lemma 12, $\sigma(\mathfrak{P}[\mathfrak{W}]\mathfrak{Q}) = \sigma(\mathfrak{P}[\mathfrak{W}\mathfrak{L}]) \geqq b(g)\sigma(\mathfrak{P}_0[\mathfrak{W}\mathfrak{L}])$

$$= b(g)\sigma(\mathfrak{P}_0[\mathfrak{W}]\mathfrak{Q}) \geqq b(g)b(h)\sigma(\mathfrak{P}_0[\mathfrak{W}]\mathfrak{Q}_0);$$

this proves the left part of the assertion. The rest follows from the inequalities $q_k \geqq q_1$ $(k = 1, \cdots, h)$, $p_k \geqq p_1$ $(k = 1, \cdots, g)$.

LEMMA 14. *Let p_1 be the first diagonal element of the reduced positive symmetric matrix $\mathfrak{P}^{(g)}$; then $\sigma(\mathfrak{P}^{-1}) < cp_1^{-1}$, where c depends only on g.*

PROOF. Since $p_k \geqq p_1$ $(k = 1, \cdots, g)$, we have $\mathfrak{P}_0 > \lambda\mathfrak{E}$, for all real $\lambda < p_1$; consequently, by Lemma 12, $\mathfrak{P} > \lambda\mathfrak{E}$ for $\lambda < bp_1$. This proves that the characteristic roots $\mu_1, \cdots, \mu_g$ of $\mathfrak{P}$ are $\geqq bp_1$. Since $\sigma(\mathfrak{P}^{-1}) = \sum_{k=1}^{g} \mu_k^{-1}$, the assertion follows, with $c = gb^{-1}$.

If $\mathfrak{P}_1^{(g)}$ is positive real symmetric, but not necessarily reduced, then we denote by $m_k(\mathfrak{P}_1)$ $(k = 1, \cdots, g)$ the diagonal elements p_k of a reduced equivalent matrix $\mathfrak{P} = \mathfrak{P}_1[\mathfrak{U}]$ with suitably chosen unimodular $\mathfrak{U}$. The value $m_1(\mathfrak{P}_1)$ can be defined independently as the minimum of the quadratic form $\mathfrak{P}_1[\mathfrak{x}]$ in the set of all integral $\mathfrak{x}^{(g)} \neq 0$.

LEMMA 15. *Let $\mathfrak{Y}^{(n)}$ be positive, $1 \leqq h \leqq n$ and let $\mathfrak{F}^{(n,h)}$ run over a complete set of non-equivalent primitive matrices; then the Dirichlet series*

$$\omega(s) = \sum_{\mathfrak{F}} |\,\mathfrak{Y}[\mathfrak{F}]\,|^{-s} \tag{27}$$

converges for $s > \frac{n}{2}$. If $m_1(\mathfrak{Y}) = y$ and $s > \frac{n}{2}$, then $\omega(s) < dy^{-hs}$, where d depends only on n and s.

PROOF. If $\mathfrak{U}^{(n)}$ is any given unimodular matrix, then also $\mathfrak{U}\mathfrak{F}$ runs over a complete set of non-equivalent primitive matrices. On the other hand, we have $|\,\mathfrak{Y}[\mathfrak{F}\mathfrak{V}]\,| = |\,\mathfrak{Y}[\mathfrak{F}]\,|$, for any unimodular $\mathfrak{V}^{(h)}$. Consequently, we may suppose that the positive symmetric matrices $\mathfrak{Y}$ and $\mathfrak{Y}[\mathfrak{F}]$ in the general term of the series (27) are both reduced, in the sense of Minkowski. Let $\mathfrak{f}_1, \cdots, \mathfrak{f}_h$ be the columns of $\mathfrak{F}$, then $\mathfrak{Y}[\mathfrak{f}_k]$ $(k = 1, \cdots, h)$ are the diagonal elements of $\mathfrak{Y}[\mathfrak{F}]$. By (26) and Lemma 12,

$$|\,\mathfrak{Y}[\mathfrak{F}]\,| > b_0^{-1}\prod_{k=1}^{h}\mathfrak{Y}[\mathfrak{f}_k] > b_0^{-1}(by)^h\prod_{k=1}^{h}[\mathfrak{f}_k]. \tag{28}$$

Let $\mathfrak{f}^{(n)}$ run over all integral columns $\neq 0$ and define

$$Z_n(s) = \sum_{\mathfrak{f}} [\mathfrak{f}]^{-s};$$

it is well known that this Dirichlet series converges for $s > \frac{n}{2}$. By (27) and (28),

we obtain the inequality

$$\omega(s) < b_0^s(by)^{-hs}Z_n(s),$$

for $s > \frac{n}{2}$, and the assertion follows readily.

5. The transformation formula

We introduce the notation

$$\eta(\mathfrak{M}) = e^{2\pi i\sigma(\mathfrak{M})},$$

for any complex square matrix $\mathfrak{M}$. Let $\mathfrak{S}$ be an integral m-rowed symmetric matrix, with the signature r, $m - r$ and abs $\mathfrak{S} = S$, let $\mathfrak{H}$ be a positive real symmetric matrix satisfying $\mathfrak{H}\mathfrak{S}^{-1}\mathfrak{H} = \mathfrak{S}$ and let $\mathfrak{Z} = \mathfrak{X} + i\mathfrak{Y}$ be a complex n-rowed symmetric matrix with positive imaginary part $\mathfrak{Y}$. We define

$$f(\mathfrak{Z}) = \sum_{\mathfrak{G}} \eta(i\mathfrak{H}[\mathfrak{G}]\mathfrak{Y} + \mathfrak{S}[\mathfrak{G}]\mathfrak{X}), \tag{29}$$

the summation extended over all integral matrices $\mathfrak{G}$ with m rows and n columns. Obviously, $f(\mathfrak{Z})$ remains invariant if $\mathfrak{H}$ is replaced by $\mathfrak{H}[\mathfrak{U}]$, for any element $\mathfrak{U}$ of $\Gamma(\mathfrak{S})$, the group of units of $\mathfrak{S}$. On the other hand, $f(\mathfrak{Z})$ is invariant under the integral modular substitutions $\mathfrak{Z} \rightarrow \mathfrak{Z}[\mathfrak{V}] + \mathfrak{T}$ with unimodular $\mathfrak{V}^{(n)}$ and integral symmetric $\mathfrak{T}^{(n)}$.

We consider a given reduced modular substitution $\mathfrak{Z} \rightarrow (\mathfrak{A}\mathfrak{Z} + \mathfrak{B})(\mathfrak{C}\mathfrak{Z} + \mathfrak{D})^{-1}$ of type h; suppose that

$$\mathfrak{A} = \begin{pmatrix} \mathfrak{A}_0 & 0 \\ 0 & \mathfrak{E} \end{pmatrix}\mathfrak{U}', \qquad \mathfrak{B} = \begin{pmatrix} \mathfrak{B}_0 & 0 \\ 0 & 0 \end{pmatrix}\mathfrak{U}^{-1}, \qquad \mathfrak{C} = \begin{pmatrix} \mathfrak{C}_0 & 0 \\ 0 & 0 \end{pmatrix}\mathfrak{U}', \qquad \mathfrak{D} = \begin{pmatrix} \mathfrak{D}_0 & 0 \\ 0 & \mathfrak{E} \end{pmatrix}\mathfrak{U}^{-1}.$$

Let $\mathfrak{G}_1$ be an integral matrix with m rows and h columns and define

$$\begin{aligned} \lambda(\mathfrak{G}_1) = {} & 2^{-\frac{1}{2}hm} e^{\pi i\frac{1}{2}h(r-\frac{1}{2}m)} S^{-h/2} \,|\, \mathfrak{C}_0 \,|^{-m/2} \\ & \cdot \sum_{\mathfrak{G}_0 \,(\mathrm{mod}\ \mathfrak{C}_0)} \eta(\tfrac{1}{4}\mathfrak{S}^{-1}[\mathfrak{G}_1]\mathfrak{A}_0\mathfrak{B}_0' - \mathfrak{S}[\mathfrak{G}_0 - \tfrac{1}{2}\mathfrak{S}^{-1}\mathfrak{G}_1\mathfrak{A}_0]\mathfrak{C}_0^{-1}\mathfrak{D}_0), \end{aligned} \tag{30}$$

where $\mathfrak{G}_0^{(m,h)}$ runs over all residue classes relative to the left ideal $(\mathfrak{C}_0)$.

The following theorem describes the behavior of $f(\mathfrak{Z})$ for any reduced modular substitution; it could easily be extended to the case of an arbitrary modular substitution, in view of Lemma 7; but we do not need this generalization.

THEOREM 3. *Let $\mathfrak{Z}^* = \mathfrak{X}^* + i\mathfrak{Y}^* = (\mathfrak{A}\mathfrak{Z} + \mathfrak{B})(\mathfrak{C}\mathfrak{Z} + \mathfrak{D})^{-1}$ be a reduced modular substitution of type h; then*

$$\begin{aligned} f(\mathfrak{Z}) = {} & |\, \mathfrak{C}\mathfrak{Z} + \mathfrak{D} \,|^{-r/2} \,|\, \mathfrak{C}\bar{\mathfrak{Z}} + \mathfrak{D} \,|^{-\frac{1}{2}(m-r)} \sum_{\mathfrak{G}_1, \mathfrak{G}_2} \lambda(\mathfrak{G}_1)\eta(i\mathfrak{H}[\tfrac{1}{2}\mathfrak{S}^{-1}\mathfrak{G}_1, \mathfrak{G}_2]\mathfrak{Y}^* \\ & + \mathfrak{S}[\tfrac{1}{2}\mathfrak{S}^{-1}\mathfrak{G}_1, \mathfrak{G}_2]\mathfrak{X}^*), \end{aligned}$$

where $\mathfrak{G}_1^{(m,h)}$ and $\mathfrak{G}_2^{(m,n-h)}$ run over all integral matrices.

PROOF. Set $\mathfrak{Z}[\mathfrak{U}] = \mathfrak{Z}_2$,

$$\mathfrak{Z}^* = \begin{pmatrix} \mathfrak{Z}_0 & \mathfrak{Z}_{01} \\ \mathfrak{Z}_{01}' & \mathfrak{Z}_1 \end{pmatrix}, \tag{31}$$

with h-rowed $\mathfrak{Z}_0$, and

$$\mathfrak{Z}_0 - \mathfrak{A}_0\mathfrak{C}_0^{-1} = \mathfrak{Z}_3, \tag{32}$$

then

$$\mathfrak{Z}_2 = \begin{pmatrix} -\mathfrak{C}_0^{-1}\mathfrak{D}_0 & 0 \\ 0 & \mathfrak{Z}_1 \end{pmatrix} - \mathfrak{Z}_3^{-1}[\mathfrak{C}_0'^{-1}, \mathfrak{Z}_{01}], \tag{33}$$

by Lemma 8. Let $\mathfrak{G}_1^{(m,h)}$, $\mathfrak{G}_2^{(m,n-h)}$ run over all integral matrices and let $\mathfrak{G}_0^{(m,h)}$ run over the same range as in (30); then $\mathfrak{G} = (\mathfrak{G}_0 + \mathfrak{G}_1\mathfrak{C}_0, \mathfrak{G}_2)$ runs exactly over all integral matrices with m rows and n columns. By (33),

$$\mathfrak{Z}_2[\mathfrak{G}'] = -\mathfrak{C}_0^{-1}\mathfrak{D}_0[\mathfrak{G}_0' + \mathfrak{C}_0'\mathfrak{G}_1'] + \mathfrak{Z}_1[\mathfrak{G}_2'] - \mathfrak{Z}_3^{-1}[\mathfrak{G}_1' + \mathfrak{C}_0'^{-1}\mathfrak{G}_0' + \mathfrak{Z}_{01}\mathfrak{G}_2'].$$

Introduce $\mathfrak{H}_1 = \frac{1}{2}(\mathfrak{S} + \mathfrak{H})$, $\mathfrak{H}_2 = \frac{1}{2}(\mathfrak{S} - \mathfrak{H})$, $\mathfrak{V} = \mathfrak{G}_0\mathfrak{C}_0^{-1} + \mathfrak{G}_2\mathfrak{Z}_{01}'$. Since $f(\mathfrak{Z}) = f(\mathfrak{Z}[\mathfrak{U}])$ and $\eta(-\mathfrak{S}[\mathfrak{G}_0 + \mathfrak{G}_1\mathfrak{C}_0]\mathfrak{C}_0^{-1}\mathfrak{D}_0] = \eta(-\mathfrak{S}[\mathfrak{G}_0]\mathfrak{C}_0^{-1}\mathfrak{D}_0)$, we obtain

$$\begin{aligned} f(\mathfrak{Z}) = f(\mathfrak{Z}_2) &= \sum_{\mathfrak{G}} \eta(\mathfrak{H}_1[\mathfrak{G}]\mathfrak{Z}_2 + \mathfrak{H}_2[\mathfrak{G}]\bar{\mathfrak{Z}}_2) \\ &= \sum_{\mathfrak{G}_0, \mathfrak{G}_2} \eta(\mathfrak{H}_1[\mathfrak{G}_2]\mathfrak{Z}_1 + \mathfrak{H}_2[\mathfrak{G}_2]\bar{\mathfrak{Z}}_1 - \mathfrak{S}[\mathfrak{G}_0]\mathfrak{C}_0^{-1}\mathfrak{D}_0) \\ &\quad \cdot \sum_{\mathfrak{G}_1} \eta(-\mathfrak{H}_1[\mathfrak{G}_1 + \mathfrak{V}]\mathfrak{Z}_3^{-1} - \mathfrak{H}_2[\mathfrak{G}_1 + \bar{\mathfrak{V}}]\bar{\mathfrak{Z}}_3^{-1}). \end{aligned} \tag{34}$$

By Lemma 3, we have

$$\mathfrak{H}_1[\mathfrak{G}_1 + \mathfrak{V}] = \mathfrak{H}_1[\mathfrak{G}_1 + \mathfrak{W}], \qquad \mathfrak{H}_2[\mathfrak{G}_1 + \bar{\mathfrak{V}}] = \mathfrak{H}_2[\mathfrak{G}_1 + \mathfrak{W}], \tag{35}$$

where

$$\mathfrak{W} = \mathfrak{G}_0\mathfrak{C}_0^{-1} + \mathfrak{S}^{-1}\mathfrak{H}_1\mathfrak{G}_2\mathfrak{Z}_{01}' + \mathfrak{S}^{-1}\mathfrak{H}_2\mathfrak{G}_2\bar{\mathfrak{Z}}_{01}'. \tag{36}$$

Let $\mathfrak{P}^{(g)}$ be a complex symmetric matrix with positive real part, let $\mathfrak{x}^{(g)}$ be a complex column, and let $\mathfrak{v}^{(g)}$ run over all integral columns; then, by the well known formula from the theory of theta functions,

$$\sum_{\mathfrak{v}} \eta(i\mathfrak{P}[\mathfrak{x} + \mathfrak{v}]) = |\,2\mathfrak{P}\,|^{-\frac{1}{2}} \sum_{\mathfrak{v}} \eta\left(\frac{i}{4}\mathfrak{P}^{-1}[\mathfrak{v}] + \mathfrak{v}'\mathfrak{x}\right).$$

We apply this formula to the inner sum in (34), using Lemmata 1, 2 and (35); then $\mathfrak{P}$ is the matrix $\mathfrak{R}$ defined in (12), with $-\mathfrak{Z}_3^{-1}$ instead of $\mathfrak{Z}$, and $\mathfrak{x}' = (\mathfrak{w}_1' \cdots \mathfrak{w}_h')$, where $\mathfrak{w}_1, \cdots, \mathfrak{w}_h$ are the columns of $\mathfrak{W}$. Since $\mathfrak{H}^{-1} = \mathfrak{H}[\mathfrak{S}^{-1}]$, we obtain

$$\begin{aligned} &\sum_{\mathfrak{G}_1} \eta(-\mathfrak{H}_1[\mathfrak{G}_1 + \mathfrak{V}]\mathfrak{Z}_3^{-1} - \mathfrak{H}_2[\mathfrak{G}_1 + \mathfrak{V}]\bar{\mathfrak{Z}}_3^{-1}) \\ &\quad = 2^{-hm/2} S^{-h/2} |\,i\mathfrak{Z}_3^{-1}\,|^{-r/2} |\,-i\bar{\mathfrak{Z}}_3^{-1}\,|^{-\frac{1}{2}(m-r)} \\ &\qquad \cdot \sum_{\mathfrak{G}_1} \eta(\mathfrak{H}_1[\tfrac{1}{2}\mathfrak{S}^{-1}\mathfrak{G}_1]\mathfrak{Z}_3 + \mathfrak{H}_2[\tfrac{1}{2}\mathfrak{S}^{-1}\mathfrak{G}_1]\bar{\mathfrak{Z}}_3 + \mathfrak{G}_1'\mathfrak{W}). \end{aligned} \tag{37}$$

By (31),

$$\eta(\mathfrak{H}_1[\mathfrak{G}_2]\mathfrak{Z}_1 + \mathfrak{H}_1[\tfrac{1}{2}\mathfrak{S}^{-1}\mathfrak{G}_1]\mathfrak{Z}_0 + \mathfrak{G}_1'\mathfrak{S}^{-1}\mathfrak{H}_1\mathfrak{G}_2\mathfrak{Z}_{01}') = \eta(\mathfrak{H}_1[\tfrac{1}{2}\mathfrak{S}^{-1}\mathfrak{G}_1, \mathfrak{G}_2]\mathfrak{Z}^*), \tag{38}$$

$$\eta(\mathfrak{H}_2[\mathfrak{G}_2]\bar{\mathfrak{Z}}_1 + \mathfrak{H}_2[\tfrac{1}{2}\mathfrak{S}^{-1}\mathfrak{G}_1]\bar{\mathfrak{Z}}_0 + \mathfrak{G}_1'\mathfrak{S}^{-1}\mathfrak{H}_2\mathfrak{G}_2\bar{\mathfrak{Z}}_{01}') = \eta(\mathfrak{H}_2[\tfrac{1}{2}\mathfrak{S}^{-1}\mathfrak{G}_1, \mathfrak{G}_2]\mathfrak{Z}^*); \tag{39}$$

by (18),

$$\begin{aligned} &\eta(\mathfrak{G}_1'\mathfrak{G}_0\mathfrak{C}_0^{-1} - \tfrac{1}{4}\mathfrak{S}^{-1}[\mathfrak{G}_1]\mathfrak{A}_0\mathfrak{C}_0^{-1} - \mathfrak{S}[\mathfrak{G}_0]\mathfrak{C}_0^{-1}\mathfrak{D}_0) \\ &\qquad = \eta(\tfrac{1}{4}\mathfrak{S}^{-1}[\mathfrak{G}_1]\mathfrak{A}_0\mathfrak{B}_0' - \mathfrak{S}[\mathfrak{G}_0 - \tfrac{1}{2}\mathfrak{S}^{-1}\mathfrak{G}_1\mathfrak{A}_0]\mathfrak{C}_0^{-1}\mathfrak{D}_0). \end{aligned} \tag{40}$$

We infer from (30), (32), (34), (36), (37), (38), (39), (40) that

$$\begin{aligned} f(\mathfrak{Z}) = {} & e^{-\pi i \frac{1}{2} h(r - \frac{1}{2}m)} \,|\, \mathfrak{C}_0 \,|^{m/2} \,|\, i\mathfrak{Z}_3^{-1} \,|^{-r/2} \,|\, -i\bar{\mathfrak{Z}}_3^{-1} \,|^{-\frac{1}{2}(m-r)} \\ & \cdot \sum_{\mathfrak{G}_1, \mathfrak{G}_2} \lambda(\mathfrak{G}_1)\eta(i\mathfrak{H}[\tfrac{1}{2}\mathfrak{S}^{-1}\mathfrak{G}_1, \mathfrak{G}_2]\mathfrak{Y}^* + \mathfrak{S}[\tfrac{1}{2}\mathfrak{S}^{-1}\mathfrak{G}_1, \mathfrak{G}_2]\mathfrak{X}^*). \end{aligned} \tag{41}$$

Since $\mathfrak{Z} = (\mathfrak{D}'\mathfrak{Z}^* - \mathfrak{B}')(\mathfrak{A}' - \mathfrak{C}'\mathfrak{Z}^*)^{-1}$, we have $(\mathfrak{C}\mathfrak{Z} + \mathfrak{D})(\mathfrak{A}' - \mathfrak{C}'\mathfrak{Z}^*) = \mathfrak{C}(\mathfrak{D}'\mathfrak{Z}^* - \mathfrak{B}') + \mathfrak{D}(\mathfrak{A}' - \mathfrak{C}'\mathfrak{Z}^*) = \mathfrak{E}$, by (18); moreover, by (19), $|\, \mathfrak{A} - \mathfrak{Z}^*\mathfrak{C} \,| = |\, \mathfrak{A}_0 - \mathfrak{Z}_0\mathfrak{C}_0 \,|$ because of $|\, \mathfrak{U} \,| = 1$; hence $|\, \mathfrak{C}\mathfrak{Z} + \mathfrak{D} \,| = |\, \mathfrak{A} - \mathfrak{Z}^*\mathfrak{C} \,|^{-1} = |\, \mathfrak{A}_0 - \mathfrak{Z}_0\mathfrak{C}_0 \,|^{-1} = |\, -\mathfrak{Z}_3\mathfrak{C}_0 \,|^{-1}$,

$$e^{-\pi i \frac{1}{2} h(r - \frac{1}{2}m)} \,|\, \mathfrak{C}_0 \,|^{m/2} \,|\, i\mathfrak{Z}_3^{-1} \,|^{-r/2} \,|\, -i\bar{\mathfrak{Z}}_3^{-1} \,|^{-\frac{1}{2}(m-r)} = |\, \mathfrak{C}\mathfrak{Z} + \mathfrak{D} \,|^{-r/2} \,|\, \mathfrak{C}\bar{\mathfrak{Z}} + \mathfrak{D} \,|^{-\frac{1}{2}(m-r)},$$

and the assertion follows from (41).

6. Estimation of $f(\mathfrak{Z})$

LEMMA 16.

$$\operatorname{abs} \lambda(\mathfrak{G}_1) \leqq 1.$$

PROOF. Set $-\frac{1}{2}\mathfrak{S}^{-1}\mathfrak{G}_1\mathfrak{A}_0 = \mathfrak{K}$, $\mathfrak{C}_0^{-1}\mathfrak{D}_0 = \mathfrak{N}_0$, and denote by G the sum on the right-hand side of (30); then

$$\operatorname{abs} \lambda(\mathfrak{G}_1) = 2^{-\frac{1}{2}hm} S^{-h/2} \,|\, \mathfrak{C}_0 \,|^{-m/2} \operatorname{abs} G, \tag{42}$$

$$\begin{aligned} \operatorname{abs} G^2 &= \sum_{\mathfrak{G}_0, \mathfrak{G}_0^* (\mathrm{mod}\ \mathfrak{C}_0)} \eta(\mathfrak{S}[\mathfrak{G}_0 + \mathfrak{G}_0^* + \mathfrak{K}]\mathfrak{N}_0 - \mathfrak{S}[\mathfrak{G}_0^* + \mathfrak{K}]\mathfrak{N}_0) \\ &= \sum_{\mathfrak{G}_0} \eta(\mathfrak{S}[\mathfrak{G}_0]\mathfrak{N}_0) \sum_{\mathfrak{G}_0^*} \eta(2\mathfrak{G}_0'\mathfrak{S}\mathfrak{G}_0^*\mathfrak{N}_0). \end{aligned} \tag{43}$$

Suppose that the matrix $2\mathfrak{S}\mathfrak{G}_0\mathfrak{C}_0^{-1}$ is integral for exactly α residue classes of $\mathfrak{G}_0$ modulo $\mathfrak{C}_0$, and let β denote the number of all residue classes. The inner sum on the right-hand side in (43) is β, whenever $2\mathfrak{S}\mathfrak{G}_0\mathfrak{N}_0$ is integral, and 0 otherwise; however, $2\mathfrak{S}\mathfrak{G}_0\mathfrak{N}_0$ is integral if and only if $2\mathfrak{S}\mathfrak{G}_0\mathfrak{C}_0^{-1}$ is integral. It follows that

$$\operatorname{abs} G^2 \leqq \alpha\beta. \tag{44}$$

In order to compute α and β, we may assume $\mathfrak{S}$ and $\mathfrak{C}_0$ to be diagonal matrices, $\mathfrak{S} = [s_1, \cdots, s_m]$ and $\mathfrak{C}_0 = [c_1, \cdots, c_h]$. We see immediately that

$$\beta = (c_1 \cdots c_h)^m = |\, \mathfrak{C}_0 \,|^m, \qquad \alpha = \prod_{k,l} (2s_k, c_l) \leqq \prod_{k=1}^{m} (2 \operatorname{abs} s_k)^h = 2^{hm} S^h, \tag{45}$$

and the assertion follows from (43), (44), (45).

We denote by $a_1, \cdots, a_{33}$ positive numbers depending only on S, m, n. For any rational symmetric $\mathfrak{R}^{(n)} = \mathfrak{C}^{-1}\mathfrak{D}$ we define

$$\gamma(\mathfrak{R}) = 2^{-\frac{1}{2}mn} e^{\pi i \frac{1}{2}n(r-\frac{1}{2}m)} S^{-n/2} \overline{|\mathfrak{R}|}^{-m} \sum_{\mathfrak{G}(\mathrm{mod}\ \mathfrak{C})} \eta(-\mathfrak{S}[\mathfrak{G}]\mathfrak{R}), \tag{46}$$

where $\mathfrak{G}$ runs over a complete system of residues modulo $\mathfrak{C}$, and

$$\varphi(\mathfrak{Z}; \mathfrak{R}) = \gamma(\mathfrak{R}) | \mathfrak{Z} + \mathfrak{R} |^{-r/2} | \bar{\mathfrak{Z}} + \mathfrak{R} |^{-\frac{1}{2}(m-r)}. \tag{47}$$

Using the definition (30) of $\lambda(\mathfrak{G}_1)$, we have $\gamma(\mathfrak{R}) = \overline{|\mathfrak{R}|}^{-m/2} \lambda(0)$.

Lemma 17. *Let $\mathfrak{Z}^* = \mathfrak{X}^* + i\mathfrak{Y}^* = (\mathfrak{A}\mathfrak{Z} + \mathfrak{B})(\mathfrak{C}\mathfrak{Z} + \mathfrak{D})^{-1}$ be a reduced modular substitution of type h and set $m_1(\mathfrak{Y}^*) = y$, $m_k(\mathfrak{H}) = h_k$ $(k = 1, \cdots, m)$; then*

$$\operatorname{abs} f(\mathfrak{Z}) < \operatorname{abs} (\mathfrak{C}\mathfrak{Z} + \mathfrak{D})^{-m/2} \prod_{k=1}^{m} (1 + a_1(h_k y)^{-n/2}) \qquad (h \leqq n),$$

$$\operatorname{abs} (f(\mathfrak{Z}) - \varphi(\mathfrak{Z}; \mathfrak{C}^{-1}\mathfrak{D})) < \operatorname{abs} (\mathfrak{C}\mathfrak{Z} + \mathfrak{D})^{-m/2} (1 + a_2(h_1 y)^{-\frac{1}{2}mn}) e^{-a_3 h_1 y} \qquad (h = n).$$

Proof. By Theorem 3, Lemma 16 and (47), we have

$$\operatorname{abs} f(\mathfrak{Z}) \leqq \operatorname{abs} (\mathfrak{C}\mathfrak{Z} + \mathfrak{D})^{-m/2} \sum_{\mathfrak{G}} \eta\left(\frac{i}{4S} \mathfrak{H}[\mathfrak{G}]\mathfrak{Y}^*\right) \qquad (h \leqq n), \tag{48}$$

where $\mathfrak{G}^{(m,n)}$ runs over all integral matrices, and

$$\operatorname{abs} (f(\mathfrak{Z}) - \varphi(\mathfrak{Z}; \mathfrak{C}^{-1}\mathfrak{D})) \leqq \operatorname{abs} (\mathfrak{C}\mathfrak{Z} + \mathfrak{D})^{-m/2} \sum_{\mathfrak{G} \neq 0} \eta\left(\frac{i}{4S} \mathfrak{H}[\mathfrak{G}]\mathfrak{Y}^*\right) \qquad (h = n). \tag{49}$$

If $\mathfrak{U}_1^{(m)}$ and $\mathfrak{U}_2^{(n)}$ are unimodular, then also $\mathfrak{U}_1\mathfrak{G}\mathfrak{U}_2'$ runs over all integral matrices or all integral matrices $\neq 0$, respectively. Therefore, in order to estimate the sums on the right-hand sides in (48) and (49), we may suppose that $\mathfrak{H}$ and $\mathfrak{Y}^*$ are reduced, in the sense of Minkowski. Then it follows from Lemma 13 that

$$\sigma(\mathfrak{H}[\mathfrak{G}]\mathfrak{Y}^*) \geqq a_4 y \sigma(\mathfrak{H}_0[\mathfrak{G}]) \geqq a_4 h_1 y \sigma([\mathfrak{G}]),$$

with $\mathfrak{H}_0 = [h_1, \cdots, h_m]$; hence

$$\eta\left(\frac{i}{4S} \mathfrak{H}[\mathfrak{G}]\mathfrak{Y}^*\right) < e^{-a_3 h_1 y} \eta\left(\frac{i}{8S} \mathfrak{H}[\mathfrak{G}]\mathfrak{Y}^*\right), \tag{50}$$

for all integral $\mathfrak{G} \neq 0$, and

$$\begin{aligned} \sum_{\mathfrak{G}} \eta\left(\frac{i}{4S} \mathfrak{H}[\mathfrak{G}]\mathfrak{Y}^*\right) &< \sum_{\mathfrak{G}} \eta\left(\frac{i}{8S} \mathfrak{H}[\mathfrak{G}]\mathfrak{Y}^*\right) \\ &< \prod_{k=1}^{m} \left(\sum_{g=-\infty}^{+\infty} e^{-a_5 h_k y g^2}\right)^n \leqq \left(\sum_{g=-\infty}^{+\infty} e^{-a_5 h_1 y g^2}\right)^{mn}, \end{aligned} \tag{51}$$

where $\mathfrak{G}$ runs over all integral matrices. Since

$$\left(\sum_{g=-\infty}^{+\infty} e^{-a_5 h_k y g^2}\right)^n < 1 + a_1(h_k y)^{-n/2},$$

the assertions follow from (48), (49), (50), (51).

7. Estimation of $\varphi(\mathfrak{Z})$

Let $\mathfrak{R}^{(n)}$ run over all rational symmetric matrices and define

$$\varphi(\mathfrak{Z}) = \sum_{\mathfrak{R}} \varphi(\mathfrak{Z}; \mathfrak{R}) = \sum_{\mathfrak{R}} \gamma(\mathfrak{R}) | \mathfrak{Z} + \mathfrak{R} |^{-r/2} | \bar{\mathfrak{Z}} + \mathfrak{R} |^{-\frac{1}{2}(m-r)}, \tag{52}$$

the expressions $\varphi(\mathfrak{Z}; \mathfrak{R})$ and $\gamma(\mathfrak{R})$ being given in (46) and (47). In virtue of a theorem of H. Braun, the Eisenstein series (52) is absolutely convergent for $2n + 2 < m$; henceforth we shall assume that this inequality is fulfilled. In order to study the behavior of $\varphi(\mathfrak{Z})$, we prove first the following

LEMMA 18. *Let $\mathfrak{K}^{(n)}$ be real symmetric, then*

$$\log \operatorname{abs} (\mathfrak{Z} + \mathfrak{K}) \leqq \log \operatorname{abs} \mathfrak{Z} + \tfrac{1}{2}\sigma(\mathfrak{Y}^{-1})\sigma^{\frac{1}{2}}(\mathfrak{K}^2).$$

PROOF. Choose the real matrix $\mathfrak{F}^{(n)}$ such that $\mathfrak{Y}[\mathfrak{F}] = \mathfrak{E}$ and $\mathfrak{X}[\mathfrak{F}] = \mathfrak{Q} = [q_1, \cdots, q_n]$, and set $\mathfrak{Q}(\mathfrak{E} + \mathfrak{Q}^2)^{-1} = \mathfrak{W}$, then $\mathfrak{Z}^{-1} = (\mathfrak{Q} + i\mathfrak{E})^{-1}[\mathfrak{F}']$, $\frac{1}{2}(\mathfrak{Z}^{-1} + \bar{\mathfrak{Z}}^{-1}) = \mathfrak{W}[\mathfrak{F}']$. If $d\mathfrak{Z}$ is real, then we get

$$\begin{aligned} d \log \operatorname{abs} \mathfrak{Z} &= \tfrac{1}{2}(d \log | \mathfrak{Z} | + d \log | \bar{\mathfrak{Z}} |) \\ &= \tfrac{1}{2}\sigma(\mathfrak{Z}^{-1}d\mathfrak{Z} + \bar{\mathfrak{Z}}^{-1}d\mathfrak{Z}) = \sigma(\mathfrak{W}[\mathfrak{F}']d\mathfrak{Z}). \end{aligned} \tag{53}$$

On the other hand, for real symmetric $\mathfrak{T}^{(m)}$ and $\mathfrak{V}^{(n)}$, we have the inequality $\operatorname{abs} \sigma(\mathfrak{T}\mathfrak{V}) \leqq \sigma^{\frac{1}{2}}(\mathfrak{T}^2)\sigma^{\frac{1}{2}}(\mathfrak{V}^2)$. Since the diagonal elements $q_k(1 + q_k^2)^{-1} (k = 1, \cdots, n)$ of the real diagonal matrix $\mathfrak{W}$ lie between $-\frac{1}{2}$ and $\frac{1}{2}$, we obtain

$$\begin{aligned} \operatorname{abs} \sigma(\mathfrak{W}[\mathfrak{F}']d\mathfrak{Z}) &\leqq \sigma^{\frac{1}{2}}(\mathfrak{W}\mathfrak{F}'\mathfrak{F}\mathfrak{W}\mathfrak{F}'\mathfrak{F})\sigma^{\frac{1}{2}}(d\mathfrak{Z}^2) \leqq \tfrac{1}{2}\sigma^{\frac{1}{2}}(\mathfrak{Y}^{-2})\sigma^{\frac{1}{2}}(d\mathfrak{Z}^2) \\ &\leqq \tfrac{1}{2}\sigma(\mathfrak{Y}^{-1})\sigma^{\frac{1}{2}}(d\mathfrak{Z}^2). \end{aligned} \tag{54}$$

The matrices $\mathfrak{Z}$ and $\mathfrak{Z} + \mathfrak{K}$ have the same imaginary part $\mathfrak{Y}$, and the assertion follows from the mean value theorem of differential calculus, by (53) and (54).

LEMMA 19. *Let $\mathfrak{Z}^* = \mathfrak{X}^* + i\mathfrak{Y}^* = (\mathfrak{A}\mathfrak{Z} + \mathfrak{B})(\mathfrak{C}\mathfrak{Z} + \mathfrak{D})^{-1}$ be a reduced modular substitution of type h, and set $m_1(\mathfrak{Y}^*) = y$; then*

$$\operatorname{abs} \varphi(\mathfrak{Z}) < a_6 \operatorname{abs} (\mathfrak{C}\mathfrak{Z} + \mathfrak{D})^{-m/2} e^{a_7 y^{-1}} \qquad (h < n),$$

$$\operatorname{abs} (\varphi(\mathfrak{Z}) - \varphi(\mathfrak{Z}; \mathfrak{C}^{-1}\mathfrak{D})) < a_8 \operatorname{abs} (\mathfrak{C}\mathfrak{Z} + \mathfrak{D})^{-m/2} e^{a_9 y^{-1}} y^{1-\frac{1}{2}m} \qquad (h = n).$$

PROOF. Define $A = \operatorname{abs} (\varphi(\mathfrak{Z}) - \varphi(\mathfrak{Z}; \mathfrak{C}^{-1}\mathfrak{D}))$, for $h = n$, and $A = \operatorname{abs} \varphi(\mathfrak{Z})$, for $h < n$. By Lemma 16 and (46) the inequality $\operatorname{abs} \gamma(\mathfrak{R}) \leqq \overline{|\mathfrak{R}|}^{-m/2}$ holds; by (52), we infer that

$$A \leqq \sum_{\mathfrak{C}_1, \mathfrak{D}_1} \operatorname{abs} (\mathfrak{C}_1 \mathfrak{Z} + \mathfrak{D}_1)^{-m/2}, \tag{55}$$

where the pair $\mathfrak{C}_1$, $\mathfrak{D}_1$ runs over all classes of coprime symmetric n-pairs, the class of $\mathfrak{C}$, $\mathfrak{D}$ excepted. Performing the inverse substitution $\mathfrak{Z} = (\mathfrak{D}'\mathfrak{Z}^* - \mathfrak{B}')(\mathfrak{A}' - \mathfrak{C}'\mathfrak{Z}^*)^{-1}$ and using the formula $(\mathfrak{A}' - \mathfrak{C}'\mathfrak{Z}^*)^{-1} = \mathfrak{C}\mathfrak{Z} + \mathfrak{D}$, we get

$$\mathfrak{C}_1\mathfrak{Z} + \mathfrak{D}_1 = (\mathfrak{C}_2\mathfrak{Z}^* + \mathfrak{D}_2)(\mathfrak{A}' - \mathfrak{C}'\mathfrak{Z}^*)^{-1} = (\mathfrak{C}_2\mathfrak{Z}^* + \mathfrak{D}_2)(\mathfrak{C}\mathfrak{Z} + \mathfrak{D}), \tag{56}$$

with

$$(\mathfrak{C}_2, \mathfrak{D}_2) = (\mathfrak{C}_1, \mathfrak{D}_1) \begin{pmatrix} \mathfrak{D}' & -\mathfrak{B}' \\ -\mathfrak{C}' & \mathfrak{A}' \end{pmatrix}.$$

Because of Lemma 6, the pair $\mathfrak{C}_2$, $\mathfrak{D}_2$ runs over all classes with exception of the 0-class. We define

$$(57)\qquad F_h = \sum_{\mathfrak{R}_0, \mathfrak{F}} \lceil \mathfrak{R}_0 \rceil^{-m/2} \operatorname{abs}(\mathfrak{Z}^*[\mathfrak{F}] + \mathfrak{R}_0)^{-m/2} \qquad (h = 1, \cdots, n),$$

where $\mathfrak{R}_0$ runs over all rational h-rowed symmetric matrices and $\mathfrak{F}^{(n,h)}$ over a complete set of non-equivalent primitive matrices. By Lemma 5, (55), (56), (57), we have

$$(58)\qquad A \leqq \operatorname{abs}(\mathfrak{C}\mathfrak{Z} + \mathfrak{D})^{-m/2} \sum_{h=1}^{n} F_h .$$

In the sum (57) we may replace $\mathfrak{F}$ by $\mathfrak{F}\mathfrak{V}$, with arbitrary unimodular $\mathfrak{V}^{(h)}$; thus we may suppose that the positive symmetric matrix $\mathfrak{Y}^*[\mathfrak{F}]$ is reduced, in the sense of Minkowski. Since $\mathfrak{Y}^*[\mathfrak{f}] \geqq m_1(\mathfrak{Y}^*) = y$, for any integral $\mathfrak{f}^{(n)} \neq 0$, it follows from Lemma 14 that

$$(59)\qquad \sigma(\mathfrak{Y}^*[\mathfrak{F}]^{-1}) < a_{10} y^{-1}.$$

Let $\mathfrak{T}^{(h)}$ be integral symmetric, and consider all real h-rowed symmetric matrices $\mathfrak{X}_h$ such that the elements of $\mathfrak{X}_h - \mathfrak{T}$ lie between 0 and 1, the value 1 excluded; they form a cube $C\,(\mathfrak{T})$ of $\frac{1}{2}h(h + 1)$ dimensions. Let dv_h be the Euclidean volume element in the space R_h of all real $\mathfrak{X}_h$, the coordinates being the $\frac{1}{2}h(h + 1)$ independent elements of $\mathfrak{X}_h$. For each $\mathfrak{R}_0$ in $C\,(\mathfrak{T})$ we obtain, by Lemma 18 and (59),

$$(60)\qquad \operatorname{abs}(\mathfrak{Z}^*[\mathfrak{F}] + \mathfrak{R}_0)^{-m/2} < e^{a_{11} y^{-1}} \int_{C(\mathfrak{T})} \operatorname{abs}(\mathfrak{Z}^*[\mathfrak{F}] + \mathfrak{X}_h)^{-m/2}\, dv_h .$$

Put $\mathfrak{R}_0 = \mathfrak{R} + \mathfrak{T}$, where $\mathfrak{R}^{(h)}$ is a given rational symmetric matrix in the unit cube $C\,(0)$, and sum (60) over all integral $\mathfrak{T}$; then

$$(61)\qquad \sum_{\mathfrak{R}_0 \equiv \mathfrak{R} \,(\mathrm{mod}\ 1)} \operatorname{abs}(\mathfrak{Z}^*[\mathfrak{F}] + \mathfrak{R}_0)^{-m/2} < e^{a_{11} y^{-1}} J(\mathfrak{F}),$$

where

$$(62)\qquad J(\mathfrak{F}) = \int_{R_h} \operatorname{abs}(\mathfrak{Z}^*[\mathfrak{F}] + \mathfrak{X}_h)^{-m/2}\, dv_h .$$

The right-hand member in (61) is independent of $\mathfrak{R}$; moreover, $\lceil \mathfrak{R}_0 \rceil = \lceil \mathfrak{R} \rceil$ for $\mathfrak{R}_0 \equiv \mathfrak{R}$ (mod 1). We multiply (61) by $\lceil \mathfrak{R} \rceil^{-m/2}$ and sum over all rational $\mathfrak{R}$ in C (0); then, in view of (57),

$$(63)\qquad F_h \leqq e^{a_{11} y^{-1}} \sum_{\mathfrak{R}\,(\mathrm{mod}\ 1)} \lceil \mathfrak{R} \rceil^{-m/2} \sum_{\mathfrak{F}} J(\mathfrak{F}).$$

Since $\frac{m}{2} > n + 1 \geqq h + 1$, the first sum in (63) converges, by Lemma 11.

We determine a real matrix $\mathfrak{K}^{(h)}$ satisfying $\mathfrak{Y}^*[\mathfrak{F}] = [\mathfrak{K}]$ and apply in the in-

tegral (62) the linear transformation $\mathfrak{X}_h \to \mathfrak{X}_h[\mathfrak{R}] - \mathfrak{X}^*[\mathfrak{F}]$, with the jacobian $|\mathfrak{R}|^{h+1} = |\mathfrak{Y}^*[\mathfrak{F}]|^{\frac{1}{2}(h+1)}$; we obtain

$$J(\mathfrak{F}) = |\mathfrak{Y}^*[\mathfrak{F}]|^{\frac{1}{2}(h-m+1)} J_h, \tag{64}$$

where

$$J_h = \int_{R_h} |\mathfrak{E} + \mathfrak{X}_h^2|^{-m/4}\, dv_h \qquad (h = 1, \cdots, n). \tag{65}$$

Because of $m - h - 1 \geqq m - n - 1 > n$, it follows from Lemma 15 and (64) that

$$\sum_{\mathfrak{F}} J(\mathfrak{F}) < a_{12}\, y^{\frac{1}{2}h(h-m+1)} J_h. \tag{66}$$

Furthermore, $\frac{h}{2}(h - m + 1) \leqq 1 - \frac{m}{2} < 0 \qquad (h = 1, \cdots, n)$, whence

$$e^{a_{11}v^{-1}}\, y^{\frac{1}{2}h(h-m+1)} < a_{13}\, e^{2a_{11}v^{-1}}\, y^{1-\frac{1}{2}m} < a_{14}\, e^{3a_{11}v^{-1}}. \tag{67}$$

In view of (58), (63), (66), (67), the assertions of the lemma will follow if we prove the convergence of the integral J_h.

We substitute $\mathfrak{X}_h = \mathfrak{W}[\mathfrak{O}]$, $\mathfrak{W} = [w_1, \cdots, w_h]$, $w_1 \geqq w_2 \geqq \cdots \geqq w_h$, with orthogonal $\mathfrak{O}$. Then $d\mathfrak{O}\mathfrak{O}'$ is a skew-symmetric matrix $\mathfrak{M}$ and $d\mathfrak{X}_h[\mathfrak{O}'] = d\mathfrak{W} + \mathfrak{W}\mathfrak{M} - \mathfrak{M}\mathfrak{W}$. Let $u_1, u_2, \cdots$ be parameters in the space of the orthogonal matrices $\mathfrak{O}^{(h)}$, and let ϕ denote the absolute value of the determinant of the $\frac{1}{2}h(h-1)$ independent elements of $\mathfrak{M}$, considered as linear functions of du_1, $du_2, \cdots$; then we have

$$dv_h = \prod_{k<l} (w_k - w_l)\, dw_1 \cdots dw_h\, \phi\, du_1\, du_2 \cdots. \tag{68}$$

If $g(\mathfrak{X}_h)$ is any integrable function of the elements of $\mathfrak{X}_h$, with the property $g(\mathfrak{X}_h) = g(\mathfrak{W})$, then

$$\int_{R_h} g(\mathfrak{X}_h)\, dv_h = a_{15} \int g(\mathfrak{W}) \prod_{k<l} (w_k - w_l)\, dw_1 \cdots dw_h, \tag{69}$$

by (68); the integration is carried over all real $w_1, \cdots, w_h$ satisfying $w_1 \geqq w_2 \geqq \cdots \geqq w_h$. It is easily proved that

$$a_{15} = \rho_h = \prod_{k=1}^{h} \frac{\pi^{k/2}}{\Gamma(k/2)};$$

however, we do not need the exact value of the finite number a_{15}. Applying (69) with $g(\mathfrak{X}_h) = |\mathfrak{E} + \mathfrak{X}_h^2|^{-m/4} = \prod_{k=1}^{h} (1 + w_k^2)^{-m/4}$ and using the inequality

$$\prod_{k<l} (w_k - w_l) \leqq \prod_{k<l} (1 + w_k^2)^{\frac{1}{2}} (1 + w_l^2)^{\frac{1}{2}} = \prod_{k=1}^{h} (1 + w_k^2)^{\frac{1}{2}(h-1)}, \tag{70}$$

we obtain the formula

$$J_h \leqq a_{15}\left(\int_{-\infty}^{+\infty}(1+w^2)^{\frac{1}{2}(h-1)-\frac{1}{2}m}\,dw\right)^h = a_{16}, \tag{71}$$

the integral being convergent, because of $\dfrac{m}{2} > n \geqq h$.

8. Approximation to $f(\mathfrak{Z})$ by $\varphi(\mathfrak{Z})$

Let Z be the space of all complex symmetric $\mathfrak{Z}^{(n)} = \mathfrak{X} + i\mathfrak{Y}$ with positive imaginary part $\mathfrak{Y}$. It is known that in Z a fundamental domain F relative to the modular group M of degree n is given by the following conditions: The inequality abs $(\mathfrak{C}_1\mathfrak{Z} + \mathfrak{D}_1) \geqq 1$ holds for all coprime symmetric n-pairs $\mathfrak{C}_1, \mathfrak{D}_1$; the matrix $\mathfrak{Y}$ is reduced, in the sense of Minkowski; all elements of $\mathfrak{X}$ lie between $-\frac{1}{2}$ and $\frac{1}{2}$. Moreover, it is known that the first diagonal element y of $\mathfrak{Y}$ satisfies the inequality $y \geqq a_{17} = \frac{1}{2}\sqrt{3}$, everywhere in F. Let G be the union of all images of F under the subgroup M_0 consisting of the integral modular substitutions $\mathfrak{Z}^* = \mathfrak{X}^* + i\mathfrak{Y}^* = \mathfrak{Z}[\mathfrak{B}] + \mathfrak{T}$, with unimodular $\mathfrak{B}^{(n)}$ and integral symmetric $\mathfrak{T}^{(n)}$. Since $\mathfrak{Y}^* = \mathfrak{Y}[\mathfrak{B}]$, we have the following

LEMMA 20. *Let $\mathfrak{Z}^* = \mathfrak{X}^* + i\mathfrak{Y}^*$ be a point of G, then $m_1(\mathfrak{Y}^*) \geqq a_{17}$.*

This lemma is the necessary tool for the generalization of the Farey dissection of the interval $0 \leqq x < 1$.

Let $K_1, K_2, \cdots$ be the classes of coprime symmetric n-pairs and let $\mathfrak{Z}^* = (\mathfrak{A}\mathfrak{Z} + \mathfrak{B})(\mathfrak{C}\mathfrak{Z} + \mathfrak{D})^{-1}$ be the reduced modular substitution, where the pair $\mathfrak{C}, \mathfrak{D}$ represents a given class $K_q(q = 1, 2, \cdots)$. Let G_q be the image of G under the inverse mapping $\mathfrak{Z} = (\mathfrak{D}'\mathfrak{Z}^* - \mathfrak{B}')(\mathfrak{A}' - \mathfrak{C}'\mathfrak{Z}^*)^{-1}$; by Lemma 7, the domains G_q do not overlap and cover the whole space Z.

We choose a positive number $\epsilon < a_{17}$ and consider the $\frac{1}{2}n(n+1)$-dimensional cube E^* defined by $\mathfrak{Z} = \mathfrak{X} + i\,\epsilon\,\mathfrak{E}$, $\mathfrak{X} = (x_{kl})$, $0 \leqq x_{kl} < 1$ $(1 \leqq k \leqq l \leqq n)$. Put

$$E^* \cap G_q = D_q, \qquad D_q - (D_1 \cup D_2 \cup \cdots \cup D_{q-1}) = E_q^*(q = 1, 2, \cdots),$$

and let E_q be the projection of E_q^* on $\mathfrak{Y} = 0$, i.e., the set consisting of the real parts $\mathfrak{X}$ of the points $\mathfrak{X} + i\,\epsilon\,\mathfrak{E}$ in E_q^*. In view of the known properties of the fundamental domain F, the sets E_q are empty for all sufficiently large q, and each E_q has a Euclidean volume in the space R_n of all real symmetric $\mathfrak{X}^{(n)}$. If E denotes the unit cube $0 \leqq x_{kl} < 1$ $(1 \leqq k \leqq l \leqq n)$ in R_n, then $E = E_1 + E_2 + \cdots$.

If K_q is the 0-class, then $G_q = G$ and E_q is empty, by Lemma 18 and the condition $\epsilon < a_{17}$. Suppose that K_q is an h-class, with $1 \leqq h \leqq n$; if

$$\mathfrak{C} = \begin{pmatrix}\mathfrak{C}_0 & 0\\ 0 & 0\end{pmatrix}\mathfrak{U}', \qquad \mathfrak{D} = \begin{pmatrix}\mathfrak{D}_0 & 0\\ 0 & \mathfrak{E}\end{pmatrix}\mathfrak{U}^{-1}$$

represent K_q, with $\mathfrak{U} = (\mathfrak{F}, \mathfrak{F}^*)$ and $\mathfrak{C}_0^{-1}\mathfrak{D}_0 = \mathfrak{R}_0$, then we write $E_q = E^*(\mathfrak{F}, \mathfrak{R}_0)$. Let $\mathfrak{R}^{(h)} = (r_{kl})$ be a given rational symmetric matrix in the unit cube $0 \leqq r_{kl}$

< 1 $(1 \leqq k \leqq l \leqq h)$, let $\mathfrak{R}_0$ run over all matrices in the residue class of $\mathfrak{R}$ modulo 1 and define

$$E_0(\mathfrak{F}, \mathfrak{R}) = \sum_{\mathfrak{R}_0 \equiv \mathfrak{R} (\mathrm{mod}\ 1)} E^*(\mathfrak{F}, \mathfrak{R}_0).$$

For any given point $\mathfrak{X}$ in $E^*(\mathfrak{F}, \mathfrak{R}_0)$, we determine the integral n-rowed symmetric matrix

$$\mathfrak{T} = \begin{pmatrix} \mathfrak{R}_0 - \mathfrak{R} & * \\ * & * \end{pmatrix}$$

such that all elements of the last $n - h$ rows in

$$\mathfrak{X}_0 = \mathfrak{X}[\mathfrak{U}] + \begin{pmatrix} \mathfrak{R} & 0 \\ 0 & 0 \end{pmatrix} + \mathfrak{T} \tag{72}$$

lie in the interval $0 \leqq x < 1$. This matrix $\mathfrak{T}$ depends on $\mathfrak{X}$, $\mathfrak{U}$, $\mathfrak{R}_0$; however, since q and E_q are bounded, $\mathfrak{T}$ belongs to a finite set. Let $E(\mathfrak{F}, \mathfrak{R})$ be the image of $E_0(\mathfrak{F}, \mathfrak{R})$ under the transformation (72). The substitution $\mathfrak{X} \rightarrow \mathfrak{X}[\mathfrak{U}] + \begin{pmatrix} \mathfrak{R} & 0 \\ 0 & 0 \end{pmatrix}$ maps the set of all residue classes of $\mathfrak{X}$ modulo 1 onto itself; consequently, $E_0(\mathfrak{F}, \mathfrak{R})$ is mapped onto $E(\mathfrak{F}, \mathfrak{R})$. On the other hand, if R_{nh} is the part of the whole space R_n defined by the conditions $0 \leqq x_{kl} < 1$ $(h < k \leqq l \leqq n)$, then $E(\mathfrak{F}, \mathfrak{R})$ is contained in R_{nh} .

Put abs $(f(\mathfrak{Z}) - \varphi(\mathfrak{Z})) = \delta(\mathfrak{Z})$. Our next aim is the estimation of the integral

$$\Delta = \int_E \delta(\mathfrak{Z})\, dv_n , \tag{73}$$

where $\mathfrak{Z} = \mathfrak{X} + i\,\epsilon\,\mathfrak{E}$. Let $\Delta(\mathfrak{F}, \mathfrak{R})$ be the integral extended over $E_0(\mathfrak{F}, \mathfrak{R})$, and define

$$\Delta_h = \sum_{\mathfrak{F}, \mathfrak{R}} \Delta(\mathfrak{F}, \mathfrak{R}) \qquad (h = 1, \cdots, n), \tag{74}$$

where $\mathfrak{R}^{(h)}$ runs over all rational symmetric matrices in the unit cube and $\mathfrak{F}^{(n,h)}$ runs over a complete set of non-equivalent primitive matrices. Plainly,

$$\Delta = \sum_{h=1}^{n} \Delta_h . \tag{75}$$

Let $\mathfrak{H}$ be the positive symmetric matrix in the definition of $f(\mathfrak{Z})$, and suppose that $m_1(\mathfrak{H}) > h_0$, with given positive h_0 . We denote by $\alpha_1 , \cdots , \alpha_8$ positive numbers depending only on h_0 , m, S.

LEMMA 21.

$$\Delta_h < \alpha_1\, \epsilon^{\frac{1}{2}h(h+1-m)} \qquad (h = 1, \cdots , n).$$

PROOF. Let $\mathfrak{X}$ be a point of $E^*(\mathfrak{F}, R_0)$, $\mathfrak{Z} = \mathfrak{X} + i\,\epsilon\,\mathfrak{E}$, and let $\mathfrak{Z}^* = (\mathfrak{A}\mathfrak{Z} + \mathfrak{B})(\mathfrak{C}\mathfrak{Z} + \mathfrak{D})^{-1}$ be the reduced modular substitution with

$$\mathfrak{C} = \begin{pmatrix} \mathfrak{C}_0 & 0 \\ 0 & 0 \end{pmatrix} \mathfrak{U}', \qquad \mathfrak{D} = \begin{pmatrix} \mathfrak{D}_0 & 0 \\ 0 & \mathfrak{E} \end{pmatrix} \mathfrak{U}^{-1}, \qquad \mathfrak{U} = (\mathfrak{F}, \mathfrak{F}^*), \qquad \mathfrak{C}_0^{-1}\mathfrak{D}_0 = \mathfrak{R}_0 ;$$

we obtain the formula

$$J_h \leqq a_{15}\left(\int_{-\infty}^{+\infty}(1+w^2)^{\frac{1}{2}(h-1)-\frac{1}{4}m}\,dw\right)^h = a_{16}, \tag{71}$$

the integral being convergent, because of $\frac{m}{2} > n \geqq h$.

8. Approximation to $f(\mathfrak{Z})$ by $\varphi(\mathfrak{Z})$

Let Z be the space of all complex symmetric $\mathfrak{Z}^{(n)} = \mathfrak{X} + i\mathfrak{Y}$ with positive imaginary part $\mathfrak{Y}$. It is known that in Z a fundamental domain F relative to the modular group M of degree n is given by the following conditions: The inequality abs $(\mathfrak{C}_1\mathfrak{Z} + \mathfrak{D}_1) \geqq 1$ holds for all coprime symmetric n-pairs $\mathfrak{C}_1, \mathfrak{D}_1$; the matrix $\mathfrak{Y}$ is reduced, in the sense of Minkowski; all elements of $\mathfrak{X}$ lie between $-\frac{1}{2}$ and $\frac{1}{2}$. Moreover, it is known that the first diagonal element y of $\mathfrak{Y}$ satisfies the inequality $y \geqq a_{17} = \frac{1}{2}\sqrt{3}$, everywhere in F. Let G be the union of all images of F under the subgroup M_0 consisting of the integral modular substitutions $\mathfrak{Z}^* = \mathfrak{X}^* + i\mathfrak{Y}^* = \mathfrak{Z}[\mathfrak{B}] + \mathfrak{T}$, with unimodular $\mathfrak{B}^{(n)}$ and integral symmetric $\mathfrak{T}^{(n)}$. Since $\mathfrak{Y}^* = \mathfrak{Y}[\mathfrak{B}]$, we have the following

LEMMA 20. *Let $\mathfrak{Z}^* = \mathfrak{X}^* + i\mathfrak{Y}^*$ be a point of G, then $m_1(\mathfrak{Y}^*) \geqq a_{17}$.*

This lemma is the necessary tool for the generalization of the Farey dissection of the interval $0 \leqq x < 1$.

Let $K_1, K_2, \cdots$ be the classes of coprime symmetric n-pairs and let $\mathfrak{Z}^* = (\mathfrak{A}\mathfrak{Z} + \mathfrak{B})(\mathfrak{C}\mathfrak{Z} + \mathfrak{D})^{-1}$ be the reduced modular substitution, where the pair $\mathfrak{C}, \mathfrak{D}$ represents a given class $K_q(q = 1, 2, \cdots)$. Let G_q be the image of G under the inverse mapping $\mathfrak{Z} = (\mathfrak{D}'\mathfrak{Z}^* - \mathfrak{B}')(\mathfrak{A}' - \mathfrak{C}'\mathfrak{Z}^*)^{-1}$; by Lemma 7, the domains G_q do not overlap and cover the whole space Z.

We choose a positive number $\epsilon < a_{17}$ and consider the $\frac{1}{2}n(n+1)$-dimensional cube E^* defined by $\mathfrak{Z} = \mathfrak{X} + i\,\epsilon\,\mathfrak{E}$, $\mathfrak{X} = (x_{kl})$, $0 \leqq x_{kl} < 1$ $(1 \leqq k \leqq l \leqq n)$. Put

$$E^* \cap G_q = D_q, \qquad D_q - (D_1 \cup D_2 \cup \cdots \cup D_{q-1}) = E_q^*(q = 1, 2, \cdots),$$

and let E_q be the projection of E_q^* on $\mathfrak{Y} = 0$, i.e., the set consisting of the real parts $\mathfrak{X}$ of the points $\mathfrak{X} + i\,\epsilon\,\mathfrak{E}$ in E_q^*. In view of the known properties of the fundamental domain F, the sets E_q are empty for all sufficiently large q, and each E_q has a Euclidean volume in the space R_n of all real symmetric $\mathfrak{X}^{(n)}$. If E denotes the unit cube $0 \leqq x_{kl} < 1$ $(1 \leqq k \leqq l \leqq n)$ in R_n, then $E = E_1 + E_2 + \cdots$.

If K_q is the 0-class, then $G_q = G$ and E_q is empty, by Lemma 18 and the condition $\epsilon < a_{17}$. Suppose that K_q is an h-class, with $1 \leqq h \leqq n$; if

$$\mathfrak{C} = \begin{pmatrix}\mathfrak{C}_0 & 0\\ 0 & 0\end{pmatrix}\mathfrak{U}', \qquad \mathfrak{D} = \begin{pmatrix}\mathfrak{D}_0 & 0\\ 0 & \mathfrak{E}\end{pmatrix}\mathfrak{U}^{-1}$$

represent K_q, with $\mathfrak{U} = (\mathfrak{F}, \mathfrak{F}^*)$ and $\mathfrak{C}_0^{-1}\mathfrak{D}_0 = \mathfrak{R}_0$, then we write $E_q = E^*(\mathfrak{F}, \mathfrak{R}_0)$. Let $\mathfrak{R}^{(h)} = (r_{kl})$ be a given rational symmetric matrix in the unit cube $0 \leqq r_{kl}$

< 1 $(1 \leqq k \leqq l \leqq h)$, let $\mathfrak{R}_0$ run over all matrices in the residue class of $\mathfrak{R}$ modulo 1 and define

$$E_0(\mathfrak{F}, \mathfrak{R}) = \sum_{\mathfrak{R}_0 \equiv \mathfrak{R} \pmod 1} E^*(\mathfrak{F}, \mathfrak{R}_0).$$

For any given point $\mathfrak{X}$ in $E^*(\mathfrak{F}, \mathfrak{R}_0)$, we determine the integral n-rowed symmetric matrix

$$\mathfrak{T} = \begin{pmatrix} \mathfrak{R}_0 - \mathfrak{R} & * \\ * & * \end{pmatrix}$$

such that all elements of the last $n - h$ rows in

$$\mathfrak{X}_0 = \mathfrak{X}[\mathfrak{U}] + \begin{pmatrix} \mathfrak{R} & 0 \\ 0 & 0 \end{pmatrix} + \mathfrak{T} \tag{72}$$

lie in the interval $0 \leqq x < 1$. This matrix $\mathfrak{T}$ depends on $\mathfrak{X}$, $\mathfrak{U}$, $\mathfrak{R}_0$; however, since q and E_q are bounded, $\mathfrak{T}$ belongs to a finite set. Let $E(\mathfrak{F}, \mathfrak{R})$ be the image of $E_0(\mathfrak{F}, \mathfrak{R})$ under the transformation (72). The substitution $\mathfrak{X} \to \mathfrak{X}[\mathfrak{U}] + \begin{pmatrix} \mathfrak{R} & 0 \\ 0 & 0 \end{pmatrix}$ maps the set of all residue classes of $\mathfrak{X}$ modulo 1 onto itself; consequently, $E_0(\mathfrak{F}, \mathfrak{R})$ is mapped onto $E(\mathfrak{F}, \mathfrak{R})$. On the other hand, if R_{nh} is the part of the whole space R_n defined by the conditions $0 \leqq x_{kl} < 1$ $(h < k \leqq l \leqq n)$, then $E(\mathfrak{F}, \mathfrak{R})$ is contained in R_{nh}.

Put abs $(f(\mathfrak{Z}) - \varphi(\mathfrak{Z})) = \delta(\mathfrak{Z})$. Our next aim is the estimation of the integral

$$\Delta = \int_E \delta(\mathfrak{Z})\, dv_n, \tag{73}$$

where $\mathfrak{Z} = \mathfrak{X} + i\,\epsilon\,\mathfrak{E}$. Let $\Delta(\mathfrak{F}, \mathfrak{R})$ be the integral extended over $E_0(\mathfrak{F}, \mathfrak{R})$, and define

$$\Delta_h = \sum_{\mathfrak{F}, \mathfrak{R}} \Delta(\mathfrak{F}, \mathfrak{R}) \qquad (h = 1, \cdots, n), \tag{74}$$

where $\mathfrak{R}^{(h)}$ runs over all rational symmetric matrices in the unit cube and $\mathfrak{F}^{(n,h)}$ runs over a complete set of non-equivalent primitive matrices. Plainly,

$$\Delta = \sum_{h=1}^{n} \Delta_h. \tag{75}$$

Let $\mathfrak{H}$ be the positive symmetric matrix in the definition of $f(\mathfrak{Z})$, and suppose that $m_1(\mathfrak{H}) > h_0$, with given positive h_0. We denote by $\alpha_1, \cdots, \alpha_8$ positive numbers depending only on h_0, m, S.

LEMMA 21.

$$\Delta_h < \alpha_1\, \epsilon^{\frac{1}{2}h(h+1-m)} \qquad (h = 1, \cdots, n).$$

PROOF. Let $\mathfrak{X}$ be a point of $E^*(\mathfrak{F}, R_0)$, $\mathfrak{Z} = \mathfrak{X} + i\,\epsilon\,\mathfrak{E}$, and let $\mathfrak{Z}^* = (\mathfrak{A}\mathfrak{Z} + \mathfrak{B})(\mathfrak{C}\mathfrak{Z} + \mathfrak{D})^{-1}$ be the reduced modular substitution with

$$\mathfrak{C} = \begin{pmatrix} \mathfrak{C}_0 & 0 \\ 0 & 0 \end{pmatrix} \mathfrak{U}', \qquad \mathfrak{D} = \begin{pmatrix} \mathfrak{D}_0 & 0 \\ 0 & \mathfrak{E} \end{pmatrix} \mathfrak{U}^{-1}, \qquad \mathfrak{U} = (\mathfrak{F}, \mathfrak{F}^*), \qquad \mathfrak{C}_0^{-1}\mathfrak{D}_0 = \mathfrak{R}_0;$$

then $\mathfrak{Z}^*$ lies in G. By Lemmata 5, 17, 19, 20,

$$\delta(\mathfrak{Z}) < \alpha_2 \overline{|\mathfrak{R}_0|}^{-m/2} \operatorname{abs}(\mathfrak{Z}[\mathfrak{F}] + \mathfrak{R}_0)^{-m/2}. \tag{76}$$

We integrate (76) over $E^*(\mathfrak{F}, \mathfrak{R}_0)$, sum over all $\mathfrak{R}_0 \equiv \mathfrak{R} \pmod 1$ and apply the linear transformation (72), with the jacobian 1. Denoting by $\mathfrak{X}_h$ the matrix of the elements in the first h rows and columns of $\mathfrak{X}_0$, we obtain $\mathfrak{X}_h = \mathfrak{X}[\mathfrak{F}] + \mathfrak{R}_0$; hence

$$\Delta(\mathfrak{F}, \mathfrak{R}) = \int_{E_0(\mathfrak{F},\mathfrak{R})} \delta(\mathfrak{Z})\, dv_n \leqq \alpha_2 \overline{|\mathfrak{R}|}^{-m/2} \int_{E(\mathfrak{F},\mathfrak{R})} \operatorname{abs}(\mathfrak{X}_h + i\epsilon[\mathfrak{F}])^{-m/2}\, dv_n. \tag{77}$$

On the other hand, $E(\mathfrak{F}, \mathfrak{R})$ is contained in R_{nh}, the product of R_h and the intervals $0 \leqq x_{kl} < 1$ $(h < k \leqq 1 \leqq n)$. It follows from (62), (64), (71) that

$$\Delta(\mathfrak{F}, \mathfrak{R}) < \alpha_3\, \epsilon^{\frac12 h(h+1-m)} \overline{|\mathfrak{R}|}^{-m/2} \,|\,[\mathfrak{F}]\,|^{\frac12(h+1-m)}.$$

Summing over $\mathfrak{F}$ and $\mathfrak{R}$, we obtain the assertion, by Lemmata 11, 15 and (74).

LEMMA 22.

$$\Delta_n < \alpha_4\, \epsilon^{\frac12 n(n+1-m)+\frac14 m-\frac12(n+1)}.$$

PROOF. Instead of (76), we infer from Lemmata 17, 19, 20 the stronger inequality

$$\delta(\mathfrak{Z}) < \alpha_5 \overline{|\mathfrak{R}_0|}^{-m/2} \operatorname{abs}(\mathfrak{Z} + \mathfrak{R}_0)^{-m/2} y^{1-\frac12 m},$$

with $y = m_1(\mathfrak{Y}^*)$, for any point $\mathfrak{X}$ of $E^*(\mathfrak{E}, \mathfrak{R}_0)$ and n-rowed $\mathfrak{R}_0$. Since

$$\begin{aligned}\mathfrak{Y}^* = \frac{1}{2i}(\mathfrak{A}\mathfrak{Z} + \mathfrak{B})(\mathfrak{C}\mathfrak{Z} + \mathfrak{D})^{-1} - \frac{1}{2i}(\bar{\mathfrak{Z}}\mathfrak{C}' + \mathfrak{D}')^{-1}(\bar{\mathfrak{Z}}\mathfrak{A}' + \mathfrak{B}')&\\ = \mathfrak{Y}[(\mathfrak{C}\mathfrak{Z} + \mathfrak{D})^{-1}] = \mathfrak{Y}[(\mathfrak{Z} + \mathfrak{R}_0)^{-1}][\mathfrak{C}^{-1}],&\end{aligned}$$

we have

$$\begin{aligned}\mathfrak{Y}^{*-1} = \mathfrak{Y}^{-1}[\mathfrak{Z} + \mathfrak{R}_0][\mathfrak{C}'] = (\mathfrak{Y}^{-1}[\mathfrak{X} + \mathfrak{R}_0] + \mathfrak{Y})[\mathfrak{C}'],&\\ \mathfrak{Y}^* = (\mathfrak{Y}^{-1}[\mathfrak{X} + \mathfrak{R}_0] + \mathfrak{Y})^{-1}[\mathfrak{C}^{-1}].&\end{aligned}$$

Because of $\mathfrak{Y} = \epsilon\,\mathfrak{E}$, the value $y = m_1(\mathfrak{Y}^*)$ is the minimum of the quadratic form $\epsilon((\mathfrak{X} + \mathfrak{R}_0)^2 + \epsilon^2\mathfrak{E})^{-1}[\mathfrak{C}^{-1}\mathfrak{x}]$ for all integral $\mathfrak{x}^{(n)} \neq 0$; the matrix $\mathfrak{C}$ is denominator of $\mathfrak{R}_0$.

Instead of (77), we obtain now the stronger inequality

$$\Delta(\mathfrak{E}, \mathfrak{R}) = \int_{E_0(\mathfrak{E},\mathfrak{R})} \delta(\mathfrak{Z})\, dv_n < \alpha_6 \overline{|\mathfrak{R}|}^{-m/2} \int_{R_n} y^{1-\frac12 m} \operatorname{abs}(\mathfrak{X}^2 + \epsilon^2\mathfrak{E})^{-m/4}\, dv_n, \tag{78}$$

where y denotes the minimum of $\epsilon(\mathfrak{X}^2 + \epsilon^2\mathfrak{E})^{-1}[\mathfrak{C}^{-1}\mathfrak{x}]$ in the set of all integral $\mathfrak{x}^{(n)} \neq 0$. We apply the transformation $\mathfrak{X} = \epsilon\,\mathfrak{W}[\mathfrak{O}]$ with orthogonal $\mathfrak{O}$ and $\mathfrak{W} = [w_1, \cdots, w_n]$, $w_1 \geqq w_2 \geqq \cdots \geqq w_n$; then

$$\epsilon^2(\mathfrak{X}^2 + \epsilon^2\mathfrak{E})^{-1}[\mathfrak{y}] = (\mathfrak{W}^2 + \mathfrak{E})^{-1}[\mathfrak{O}\mathfrak{y}] \geqq (w_1^2 + 1)^{-1}[\mathfrak{O}\mathfrak{y}] = (w_1^2 + 1)^{-1}[\mathfrak{y}],$$

for real $\mathfrak{y}^{(n)}$. Since $\overline{|\mathfrak{R}|}\,\mathfrak{C}^{-1}$ is integral, we have $[\mathfrak{C}^{-1}\mathfrak{x}] \geqq \overline{|\mathfrak{R}|}^{-2}$, for all integral $\mathfrak{x} \neq 0$. Consequently, $y \geqq (w_1^2 + 1)^{-1}\,\epsilon^{-1}\,\overline{|\mathfrak{R}|}^{-2}$.

Define

$$a_{17}^{-\frac{1}{2}}\,\epsilon^{-\frac{1}{2}}\,\overline{|\mathfrak{R}|}^{-1} = \vartheta, \qquad a_{17}\max\,(1, \vartheta^2(w_1^2+1)^{-1}) = y_0\,;$$

then $y \geqq y_0$, in view of Lemma 20. Replacing y by y_0 in (78), we obtain an integrand depending only on $w_1, \cdots, w_n$; thus we can apply (69). Because of (70),

$$\Delta(\mathfrak{E}, \mathfrak{R}) < \alpha_6\,\overline{\quad\;}|^{-m/2}\,\epsilon^{\frac{1}{2}n(n+1-m)}\int_{-\infty}^{+\infty} y^{1-\frac{1}{2}m}(1+w_1^2)^{\frac{1}{2}(n-1)-\frac{1}{4}m}\,dw_1\,.$$

Denote the integral by J. If $\vartheta \leqq 1$, then $y_0 = a_{17}$ and $J = a_{18}$. If $\vartheta > 1$, then

$$J < a_{19}\int_{\vartheta}^{\infty} w^{n-1-\frac{1}{2}m}\,dw + a_{20}\,\vartheta^{2-m}\int_{1}^{\vartheta} w^{n+\frac{1}{2}m-3}\,dw < a_{21}\,\vartheta^{n-\frac{1}{2}m}.$$

Hence in both cases

$$J < a_{22}\min\,(1, \epsilon^{\frac{1}{4}m-\frac{1}{2}n}\,\overline{|\mathfrak{R}|}^{\frac{1}{2}m-n}).$$

Summing over all rational symmetric $\mathfrak{R}^{(n)}$ in the unit cube, we get

$$\Delta_n = \sum_{\mathfrak{R}}\Delta(\mathfrak{E}, \mathfrak{R}) < \alpha_7\,\epsilon^{\frac{1}{2}n(n+1-m)}\Big(\epsilon^{\frac{1}{4}m-\frac{1}{2}n}\sum_{\overline{|\mathfrak{R}|}<\epsilon^{-\frac{1}{2}}}\overline{|\mathfrak{R}|}^{-n} + \sum_{\overline{|\mathfrak{R}|}\geqq\epsilon^{-\frac{1}{2}}}\overline{|\mathfrak{R}|}^{-m/2}\Big),$$

and the assertion follows from Lemma 11, with $u = \epsilon^{-\frac{1}{2}}$ and $s = \dfrac{m}{2} - n > 1$.

If $1 \leqq h \leqq n-1$, then $\dfrac{h}{2}(h+1-m) \geqq \dfrac{n-1}{2}(n-m) > \dfrac{n}{2}(n+1-m) + \dfrac{m}{4} - \dfrac{n+1}{2}$. We estimate Δ_h for $h = 1, \cdots, n-1$ by Lemma 21 and for $h = n$ by Lemma 22. By (73), (75), we obtain the important

LEMMA 23. *Let $\mathfrak{Z} = \mathfrak{X} + i\,\epsilon\,\mathfrak{E}$, with $0 < \epsilon < a_{17}$, and suppose that $m_1(\mathfrak{H}) > h_0 > 0$, then*

$$\int_E \text{abs}\,(f(\mathfrak{Z}) - \varphi(\mathfrak{Z}))\,dv_n < \alpha_8\,\epsilon^{\frac{1}{2}n(n+1-m)+\frac{1}{4}m-\frac{1}{2}(n+1)},$$

where α_8 depends only on h_0, m, S.

9. Integration over $\mathfrak{X}$

Let $\mathfrak{T}^{(n)}$ be integral symmetric and define

$$T(\epsilon) = \sum_{\mathfrak{S}[\mathfrak{G}]=\mathfrak{T}} e^{-\frac{1}{2}\pi\epsilon\sigma(\mathfrak{H}[\mathfrak{G}])}; \tag{79}$$

the summation is extended over all integral solutions $\mathfrak{G}$ of $\mathfrak{S}[\mathfrak{G}] = \mathfrak{T}$ and ϵ is a positive number. Multiplying (29) by $\eta(-\mathfrak{T}\mathfrak{X})$ and integrating over $\mathfrak{X}$ in the unit cube E, we obtain

$$T(4\epsilon) = \int_E f(\mathfrak{Z})\eta(-\mathfrak{T}\mathfrak{X})\,dv_n, \tag{80}$$

with $\mathfrak{Z} = \mathfrak{X} + i\,\epsilon\,\mathfrak{E}$. Set $\frac{n}{2}(n + 1 - m) = j$, and apply Lemma 23; since $n + 1 < 2m$, we infer that

$$T(4\epsilon) = \int_E \varphi(\mathfrak{Z})\eta(-\mathfrak{T}\mathfrak{X})\,dv_n + o(\epsilon^j); \tag{81}$$

the sign o refers to the passage to the limit $\epsilon \to 0$ and it holds uniformly in $\mathfrak{H}$ for $m_1(\mathfrak{H}) > h_0$, where h_0 is any given positive number.

In view of the definition (52) of $\varphi(\mathfrak{Z})$, we have

$$\int_E \varphi(\mathfrak{Z})\eta(-\mathfrak{T}\mathfrak{X})\,dv_n = \sum_{\mathfrak{R}\,(\mathrm{mod}\,1)} \gamma(\mathfrak{R})\eta(\mathfrak{T}\mathfrak{R}) \cdot \int_{R_n} |\mathfrak{Z}|^{-r/2}\,|\overline{\mathfrak{Z}}|^{-\frac{1}{2}(m-r)}\,\eta(-\mathfrak{T}\mathfrak{X})\,dv_n, \tag{82}$$

where $\mathfrak{R}$ runs over all rational n-rowed symmetric matrices modulo 1; the interchange of integration and summation was allowed, the proof of Lemma 19 showing the uniform convergence of the series $\varphi(\mathfrak{Z})$ in E. Also,

$$\int_{R_n} |\mathfrak{Z}|^{-r/2}\,|\overline{\mathfrak{Z}}|^{-\frac{1}{2}(m-r)}\,\eta(-\mathfrak{T}\mathfrak{X})\,dv_n = \epsilon^j e^{\pi i \frac{1}{2} n(\frac{1}{2}m - r)} \int_{R_n} |\mathfrak{E} - i\mathfrak{X}|^{-r/2}\,|\mathfrak{E} + i\mathfrak{X}|^{-\frac{1}{2}(m-r)}\,\eta(-\epsilon\mathfrak{T}\mathfrak{X})\,dv_n. \tag{83}$$

The integral on the right-hand side is absolutely convergent, by (65), (71), and the absolute value of the integrand does not depend on ϵ; thus the integral converges uniformly with respect to ϵ, and we may interchange the integration and the passage to the limit $\epsilon \to 0$. Define

$$J_n(\alpha, \beta) = \int_{R_n} |\mathfrak{E} - i\mathfrak{X}|^{-\alpha}\,|\mathfrak{E} + i\mathfrak{X}|^{-\beta}\,dv_n \qquad (\alpha + \beta > n),$$

then it follows from (81), (82), (83) that

$$\epsilon^{-j}\,T(4\epsilon) = e^{\pi i \frac{1}{2} n(\frac{1}{2}m - r)} J_n\left(\frac{r}{2}, \frac{m-r}{2}\right) \sum_{\mathfrak{R}\,(\mathrm{mod}\,1)} \gamma(\mathfrak{R})\eta(\mathfrak{T}\mathfrak{R}) + o(1), \tag{84}$$

uniformly in $\mathfrak{H}$ for $m_1(\mathfrak{H}) > h_0 > 0$.

Lemma 24. *Let α, β be real and $\alpha + \beta > n$, then*

$$J_n(\alpha, \beta) = 2^{-n(\alpha+\beta)}(4\pi)^{\frac{1}{2}n(n+3)} \prod_{k=1}^{n-1} \frac{\Gamma\left(\alpha + \beta - \frac{k+n+1}{2}\right)}{\Gamma\left(\alpha - \frac{k}{2}\right)\Gamma\left(\beta - \frac{k}{2}\right)}. \tag{85}$$

Proof. Let λ and μ be complex constants with positive real parts. It is well known that

$$\int_{-\infty}^{+\infty} (\lambda - ix)^{-\alpha}(\mu + ix)^{-\beta}\,dx = 2\pi(\mu + \lambda)^{1-\alpha-\beta}\,\frac{\Gamma(\alpha + \beta - 1)}{\Gamma(\alpha)\Gamma(\beta)} \qquad (\alpha + \beta > 1). \tag{86}$$

This proves the lemma in case $n = 1$. Suppose that $n > 1$ and apply induction. Setting

$$\mathfrak{E} + i\mathfrak{X} = \begin{pmatrix} 1 + ix & i\mathfrak{x}' \\ i\mathfrak{x} & \mathfrak{E}_1 + i\mathfrak{X}_1 \end{pmatrix},$$

with $\mathfrak{E}_1 = \mathfrak{E}^{(n-1)}$, we have

$$|\mathfrak{E} + i\,\mathfrak{X}| = |\mathfrak{E}_1 + i\mathfrak{X}_1|\,(1 + ix + (\mathfrak{E}_1 + i\mathfrak{X}_1)^{-1}[\mathfrak{x}]).$$

We use (86) with $\mu = 1 + (\mathfrak{E}_1 + i\mathfrak{X}_1)^{-1}[\mathfrak{x}]$ and $\lambda = 1 + (\mathfrak{E}_1 - i\mathfrak{X}_1)^{-1}[\mathfrak{x}]$; because of $\lambda + \mu = 2 + 2(\mathfrak{E}_1 + \mathfrak{X}_1^2)^{-1}[\mathfrak{x}]$, we get

$$\begin{aligned} J_n(\alpha, \beta) = 2^{2-\alpha-\beta}\,\pi\,\frac{\Gamma(\alpha + \beta - 1)}{\Gamma(\alpha)\Gamma(\beta)} \int_{R_{n-1}} |\mathfrak{E}_1 - i\mathfrak{X}_1|^{-\alpha}\,|\mathfrak{E}_1 + i\mathfrak{X}_1|^{-\beta} \\ \cdot\left(\int_{-\infty}^{+\infty} (1 + (\mathfrak{E}_1 + \mathfrak{X}_1^2)^{-1}[\mathfrak{x}])^{1-\alpha-\beta}\{d\mathfrak{x}\}\right) dv_{n-1}, \end{aligned} \tag{87}$$

where $\{d\mathfrak{x}\}$ denotes the $(n - 1)$-dimensional Euclidean volume element in the $\mathfrak{x}$-space. Also, if $\mathfrak{P}^{(n)}$ is positive symmetric,

$$\begin{aligned} \int_{-\infty}^{+\infty} (1 + \mathfrak{P}[\mathfrak{x}])^{1-\alpha-\beta}\{d\mathfrak{x}\} &= \frac{\pi^{\frac{1}{2}(n-1)}}{\Gamma\left(\frac{n-1}{2}\right)}\,|\mathfrak{P}|^{-\frac{1}{2}} \int_0^{\infty} x^{\frac{1}{2}(n-1)-1}(1 + x)^{1-\alpha-\beta}\,dx \\ &= \pi^{\frac{1}{2}(n-1)}\,\frac{\Gamma\left(\alpha + \beta - \frac{n+1}{2}\right)}{\Gamma(\alpha + \beta - 1)}\,|\mathfrak{P}|^{-\frac{1}{2}}, \end{aligned}$$

provided $\alpha + \beta > \dfrac{n+1}{2}$; choose $\mathfrak{P} = (\mathfrak{E}_1 + \mathfrak{X}_1^2)^{-1}$, then (87) leads to the recursion formula

$$J_n(\alpha, \beta) = 2^{2-\alpha-\beta}\,\pi^{\frac{1}{2}(n+1)}\,\frac{\Gamma\left(\alpha + \beta - \frac{n+1}{2}\right)}{\Gamma(\alpha)\Gamma(\beta)}\,J_{n-1}(\alpha - \tfrac{1}{2}, \beta - \tfrac{1}{2}).$$

Applying (85) with $\alpha - \frac{1}{2}, \beta - \frac{1}{2}, n - 1$ instead of α, β, n, we obtain the assertion.

Let $q = q_1, q_2, \cdots$ run over a sequence Q of positive integers such that $t \mid q_k$ for any positive integer t and all sufficiently large k; e.g., the sequence $q_k = k!$ has the required property. Let $\alpha_p(\mathfrak{S}, \mathfrak{T})$ be the p-adic density defined in the introduction and set

$$\lambda(\mathfrak{S}, \mathfrak{T}) = \prod_p \alpha_p(\mathfrak{S}, \mathfrak{T}),$$

the product extended over all primes p.

LEMMA 25. *Let q run over a sequence Q, then*

$$\begin{aligned} 2^{\frac{1}{2}n(m-n+1)}\,S^{m/2}\,e^{\pi i\frac{1}{2}n(\frac{1}{2}m-r)} \sum_{\mathfrak{R}\,(\mathrm{mod}\,1)} \gamma(\mathfrak{R})\eta(\mathfrak{T}\mathfrak{R}) \\ = \lim_{q\to\infty} q^{\frac{1}{2}n(n+1)-mn}\,A_q(\mathfrak{S}, \mathfrak{T}) = \lambda(\mathfrak{S}, \mathfrak{T}), \end{aligned} \tag{88}$$

and the limit exists uniformly with respect to $\mathfrak{T}$.

PROOF. We say that a symmetric matrix $\mathfrak{W}$ is semi-integral, whenever $2\mathfrak{W}$ is integral with even diagonal elements. Plainly, if $\mathfrak{R}$ is any rational symmetric matrix, then $q\mathfrak{R} = \mathfrak{W}$ is semi-integral for all sufficiently large q in the sequence Q. Denote the left-hand member in (88) by B, and use the definition of $\gamma(\mathfrak{R})$ in (46). By Lemmata 11 and 16, the sum is absolutely convergent, and uniformly with respect to $\mathfrak{T}$, whence

$$(89)\qquad 2^{\frac{1}{2}n(n-1)} B = \lim_{q\to\infty} \sum_{\mathfrak{W}\,(\mathrm{mod}\, q)} \mathrm{abs}\, \mathfrak{C}^{-m} \sum_{\mathfrak{G}\,(\mathrm{mod}\,\mathfrak{C})} \eta(q^{-1}(\mathfrak{T} - \mathfrak{S}[\mathfrak{G}])\mathfrak{W}),$$

uniformly in $\mathfrak{T}$, where $\mathfrak{C}$ is denominator of $q^{-1}\mathfrak{W}$ and $\mathfrak{W}^{(n)}$ runs over a complete set of semi-integral symmetric matrices modulo q. Since the number of integral residue classes $\mathfrak{G}^{(m,n)}$ modulo $\mathfrak{C}$ is abs $\mathfrak{C}^m$, it follows that the general term of the outer sum in (89) is not changed, if we replace $\mathfrak{C}$ by any integral non-singular $\mathfrak{C}_1^{(n)}$ such that the value of $\sigma(\mathfrak{S}[\mathfrak{G}]\mathfrak{W})$ modulo q depends only on the residue class of $\mathfrak{G}$ modulo $\mathfrak{C}_1$, for fixed $\mathfrak{W}$. In particular, we may choose $\mathfrak{C}_1 = q\mathfrak{E}$; then $\mathfrak{C}_1$ is independent of $\mathfrak{W}$, and the two summations in (89) can be interchanged.

For any integral symmetric $\mathfrak{M}^{(n)}$, we have

$$\sum_{\mathfrak{W}\,(\mathrm{mod}\, q)} \eta(q^{-1}\mathfrak{M}\mathfrak{W}) = 0,$$

except when all elements of $\mathfrak{M}$ are divisible by q; in the latter case the sum is equal to the number of semi-integral $\mathfrak{W}$ modulo q, namely $2^{\frac{1}{2}n(n-1)}q^{\frac{1}{2}n(n+1)}$. Consequently,

$$B = \lim_{q\to\infty} q^{\frac{1}{2}n(n+1)-mn} A_q(\mathfrak{S}, \mathfrak{T}),$$

uniformly in $\mathfrak{T}$, where $A_q(\mathfrak{S}, \mathfrak{T})$ is the number of modulo q incongruent integral solutions $\mathfrak{G}$ of $\mathfrak{S}[\mathfrak{G}] \equiv \mathfrak{T} \pmod q$. This proves the left part of the assertion.

On the other hand let the sequence $p_1 = 2, p_2, p_3, \cdots$ consist of all different prime numbers p, and let $q = q_1q_2\cdots$, $q_k = p_k^{b_k}$ $(k = 1, 2, \cdots)$ with non-negative integral b_k. Of course, $b_k = 0$ for all sufficiently large k; suppose that this is true for all $k > \nu$, the positive integer ν depending on q. Let $\mathfrak{W}_k^{(n)}$ $(k = 2, \cdots, \nu)$ run over all integral residue classes modulo q_k and let $\mathfrak{W}_1^{(n)}$ run over all semi-integral residue classes modulo q_1, then $\mathfrak{W} = q(q_1^{-1}\mathfrak{W}_1 + \cdots + q_\nu^{-1}\mathfrak{W}_\nu)$ runs exactly over all semi-integral residue classes modulo q. Also, if $\mathfrak{G}_k^{(m,n)}(k = 1, \cdots, \nu)$ runs over all integral residue classes modulo q_k, then $\mathfrak{G} = q(q_1^{-1}\mathfrak{G}_1 + \cdots + q_\nu^{-1}\mathfrak{G}_\nu)$ does the same modulo q. Since

$$\eta(q^{-1}(\mathfrak{T} - \mathfrak{S}[\mathfrak{G}])\mathfrak{W}) = \prod_{k=1}^{\nu} \eta(q_k^{-1}(\mathfrak{T} - \mathfrak{S}[q_k^{-1}q\mathfrak{G}_k])\mathfrak{W}_k)$$

and the matrices $\mathfrak{G}_k$, $q_k^{-1}q\mathfrak{G}_k$ run at the same time over all integral residue classes modulo q_k, it follows that the sum in (89) is equal to the product of the ν sums

$$s_k = \sum_{\mathfrak{W}_k} q_k^{-m} \sum_{\mathfrak{G}_k\,(\mathrm{mod}\, q_k)} \eta(q_k^{-1}(\mathfrak{T} - \mathfrak{S}[\mathfrak{G}_k])\mathfrak{W}_k) \qquad (k = 1, \cdots, \nu)$$

and that the terms of the outer sum in (89) are obtained by multiplication of the outer terms in the sums $s_1, \cdots, s_\nu$. In view of the absolute convergence of the sum in (88), we infer that $\lim_{b_k \to \infty} s_k$ exists for each $k = 1, 2, \cdots$, and that

$$2^{\frac{1}{2}n(n-1)} B = \prod_{k=1}^{\infty} \lim_{b_k \to \infty} s_k .$$

Also,

$$s_1 = 2^{\frac{1}{2}n(n-1)} q_1^{\frac{1}{2}n(n+1)-mn} A_{q_1}(\mathfrak{S}, \mathfrak{T}), \quad s_k = q_k^{\frac{1}{2}n(n+1)-mn} A_{q_k}(\mathfrak{S}, \mathfrak{T}) \quad (k = 2, 3, \cdots);$$

consequently the densities $\alpha_p(\mathfrak{S}, \mathfrak{T})$ exist and

$$B = \prod_p \alpha_p(\mathfrak{S}, \mathfrak{T}) = \lambda(\mathfrak{S}, \mathfrak{T}).$$

This proves the right part of the assertion.

By (84) and Lemmata 24, 25, we obtain

$$\epsilon^{-j}\, T(\epsilon) = \pi^{\frac{1}{2}n(n+3)} S^{-n/2} \lambda(\mathfrak{S}, \mathfrak{T}) \prod_{k=0}^{n-1} \frac{\Gamma\left(\frac{m-n-k-1}{2}\right)}{\Gamma\left(\frac{r-k}{2}\right)\Gamma\left(\frac{m-r-k}{2}\right)} + o(1). \tag{90}$$

Set

$$\rho_l^{-1} = \prod_{k=1}^{l} \pi^{-k/2}\, \Gamma\left(\frac{k}{2}\right) \qquad (l = 1, 2, \cdots), \tag{91}$$

$\rho_0 = 1$ and $\rho_l^{-1} = 0$ for $l < 0$, then (79) and (90) imply

THEOREM 4. *If $2n + 2 < m$, then*

$$\lim_{\epsilon \to 0} \epsilon^{\frac{1}{2}n(m-n-1)} \sum_{\mathfrak{S}[\mathfrak{G}]=\mathfrak{T}} e^{-\frac{1}{2}\pi\epsilon\sigma(\mathfrak{H}[\mathfrak{G}])} = \frac{\rho_r\, \rho_{m-r}\, \rho_{m-2n-1}}{\rho_{r-n}\, \rho_{m-r-n}\, \rho_{m-n-1}} S^{-n/2} \prod_p \alpha_p(\mathfrak{S}, \mathfrak{T}),$$

uniformly for $m_1(\mathfrak{H}) > h_0 > 0$.

For the later proof of Theorem 1, we need the following estimate of $T(\epsilon)$ which holds uniformly in the whole $\mathfrak{H}$-space.

THEOREM 5. *Suppose that $2n + 2 < m$ and $0 < \epsilon < 1$, then*

$$\epsilon^{\frac{1}{2}n(m-n-1)} \sum_{\mathfrak{S}[\mathfrak{G}]=\mathfrak{T}} e^{-\frac{1}{2}\pi\epsilon\sigma(\mathfrak{H}[\mathfrak{G}])} < a_{23} \prod_{k=1}^{m} (1 + h_k^{-n/2}),$$

where $h_k = m_k(\mathfrak{H})$ $(k = 1, \cdots, m)$ and a_{23} depends only on m and S.

PROOF. We have to repeat the argument leading to Lemma 21. Suppose that $0 < \epsilon < a_{17}$ and put

$$\int_E \operatorname{abs} f(\mathfrak{Z})\, dv_n = \sum_{h=1}^{n} \Delta_h^*, \tag{92}$$

with $\mathfrak{Z} = \mathfrak{X} + i\,\epsilon\,\mathfrak{E}$, where Δ_h^* is the integral extended over the union of all domains $E_0(\mathfrak{F}, \mathfrak{R})$ with h-rowed $\mathfrak{R}$. Instead of (76), we use the inequality

$$\operatorname{abs} f(\mathfrak{Z}) < a_{24}\, |\mathfrak{R}_0|^{-m/2} \operatorname{abs} (\mathfrak{Z}[\mathfrak{F}] + \mathfrak{R}_0)^{-m/2} \prod_{k=1}^{m} (1 + h_k^{-n/2}) \tag{93}$$

following from Lemmata 5, 17, 20, for all points $\mathfrak{X}$ in $E^*(\mathfrak{F}, \mathfrak{R}_0)$. If we replace the factor a_2 on the right-hand side in (76) by $a_{24}\prod_{k=1}^{m}(1 + h_k^{-n/2})$, then we get the right-hand side in (93); we remark that the latter factor does not depend on $\mathfrak{Z}$. Otherwise, the proof remains the same, and we infer that

$$\Delta_h^* < a_{25}\,\epsilon^{\frac{1}{2}h(h+1-m)} \prod_{k=1}^{m} (1 + h_k^{-n/2}). \tag{94}$$

On the other hand, $h(h + 1 - m) \geqq n(n + 1 - m)$, for $h = 1, \cdots, n$, and

$$T(4\epsilon) \leqq \int_E \operatorname{abs} f(\mathfrak{Z})\, dv_n\,, \tag{95}$$

by (80). Because of $a_{17} = \frac{1}{2}\sqrt{3} > \frac{1}{4}$, the assertion follows from (79), (92), (94), (95), if we replace ϵ by $\frac{\epsilon}{4}$.

10. Integration over $\mathfrak{H}$

If $\mathfrak{X} = (x_{kl})$ is a real matrix with independently variable elements, we define $\{d\mathfrak{X}\} = \prod_{k,l} dx_{kl}$; analogously, for symmetric $\mathfrak{X}$, we define $\{d\mathfrak{X}\} = \prod_{k \leqq l} dx_{kl}$. For any complex square matrix designated by a capital Gothic letter, we denote the absolute value of the determinant by the corresponding capital Roman letter; e.g., abs $\mathfrak{T} = T$.

Let $\mathfrak{B}^{(\mu)}$ be a non-singular real symmetric matrix of signature α, $\mu - \alpha$, and let $\mathfrak{F}^{(\mu,\nu)}$ be a real matrix such that $\mathfrak{B}[\mathfrak{F}] = \mathfrak{T}$ has the signature α, $\nu - \alpha$. If $\mathfrak{X}^{(\mu,\lambda)}$ is a variable real matrix, then the formula

$$\mathfrak{B}[\mathfrak{F}, \mathfrak{X}] = \mathfrak{W} = \begin{pmatrix} \mathfrak{T} & \mathfrak{Q} \\ \mathfrak{Q}' & \mathfrak{R} \end{pmatrix}$$

with $\mathfrak{Q} = \mathfrak{F}'\mathfrak{B}\mathfrak{X}$, $\mathfrak{R} = \mathfrak{B}[\mathfrak{X}]$, defines a mapping of the $\mathfrak{X}$-space into the $\mathfrak{Q}$, $\mathfrak{R}$-space. Suppose that $\mathfrak{W}$ is non-singular, then $\nu + \lambda \leqq \mu$, and the signature of $\mathfrak{W}$ is α, $\nu + \lambda - \alpha$.

Lemma 26. *Let D be any domain in the $\mathfrak{Q}$, $\mathfrak{R}$-space such that $\mathfrak{W}$ has the signature α, $\nu + \lambda - \alpha$, let D_0 be the domain in the $\mathfrak{X}$-space which is mapped into D, and define $T = 1$ for $\nu = 0$; then*

$$\int_{D_0} g(\mathfrak{W})\{d\mathfrak{X}\} = \frac{\rho_{\mu-\nu}}{\rho_{\mu-\nu-\lambda}}\,(VT)^{-\lambda/2} \int_D g(\mathfrak{W})(TW^{-1})^{\frac{1}{2}(\lambda+\nu-\mu+1)}\{d\mathfrak{Q}\}\{d\mathfrak{R}\}, \tag{96}$$

for any integrable function $g(\mathfrak{W})$ of the elements of $\mathfrak{W}$.

Proof. In view of the mean value theorem, it suffices to prove the assertion for $g(\mathfrak{W}) = 1$. We consider first the particular case $\lambda = 1$, $\nu = 0$; then $\alpha = 0$,

$-\mathfrak{B}$ is a positive symmetric matrix $\mathfrak{P}^{(\mu)}$ and $\mathfrak{W} = \mathfrak{R}$ is a negative number $r = -p$. The volume of the ellipsoid $\mathfrak{P}[\mathfrak{x}] \leqq p$ is

$$s(p) = \frac{\pi^{\mu/2}}{\Gamma\left(\frac{\mu}{2} + 1\right)} \, | \mathfrak{P} |^{-\frac{1}{2}} p^{\mu/2};$$

moreover, $\pi^{\mu/2}/\Gamma\left(\frac{\mu}{2}\right) = \rho_\mu/\rho_{\mu-1}$, by (91); hence

$$\text{(97)} \qquad \int_{\mathfrak{P}[\mathfrak{x}] \leqq a} \{d\mathfrak{x}\} = \int_0^a \frac{ds(p)}{dp}\, dp = \frac{\rho_\mu}{\rho_{\mu-1}} \, | \mathfrak{P} |^{-\frac{1}{2}} \int_0^a p^{\frac{1}{2}\mu-1}\, dp.$$

This proves (96) for $\lambda = 1$, $\nu = 0$.

Suppose next that $\lambda = 1$, $\nu > 0$; then $\mathfrak{Q} = \mathfrak{q}$ is a column of ν elements and $\mathfrak{R} = r$ is again a number. We choose a real matrix $\mathfrak{F}_1^{(\mu,\mu-\nu)}$ such that $(\mathfrak{F}, \mathfrak{F}_1) = \mathfrak{M}$ is non-singular and set

$$\text{(98)} \qquad \mathfrak{M}^{-1}(\mathfrak{F}, \mathfrak{x}) = \begin{pmatrix} \mathfrak{E}^{(\nu)} & \mathfrak{y}_1 \\ 0 & \mathfrak{y}_2 \end{pmatrix}, \qquad \mathfrak{B}[\mathfrak{M}] = \begin{pmatrix} \mathfrak{T} & \mathfrak{Q}_0 \\ \mathfrak{Q}_0' & \mathfrak{R}_0 \end{pmatrix} = \mathfrak{W}_0,$$

$$\mathfrak{R}_0 - \mathfrak{T}^{-1}[\mathfrak{Q}_0] = -\mathfrak{P}_0, \qquad r - \mathfrak{T}^{-1}[\mathfrak{q}] = -p;$$

then

$$\text{(99)} \qquad \mathfrak{W}_0 = \begin{pmatrix} \mathfrak{T} & 0 \\ 0 & -\mathfrak{P}_0 \end{pmatrix} \begin{bmatrix} \mathfrak{E} & \mathfrak{T}^{-1}\mathfrak{Q}_0 \\ 0 & \mathfrak{E} \end{bmatrix}, \qquad \mathfrak{W} = \begin{pmatrix} \mathfrak{T} & 0 \\ 0 & -p \end{pmatrix} \begin{bmatrix} \mathfrak{E} & \mathfrak{T}^{-1}\mathfrak{q} \\ 0 & 1 \end{bmatrix},$$

$$\text{(100)} \qquad \mathfrak{T}^{-1}\mathfrak{q} = \mathfrak{y}_1 + \mathfrak{T}^{-1}\mathfrak{Q}_0\mathfrak{y}_2, \qquad \mathfrak{P}_0[\mathfrak{y}_2] = p.$$

Since $\mathfrak{B}$ and $\mathfrak{T}$ have the signatures α, $\mu - \alpha$ and α, $\nu - \alpha$, it follows that $\mathfrak{P}_0$ and p are positive. Instead of $\mathfrak{x}$, we introduce the variables $\mathfrak{q}$, $\mathfrak{y}_2$ into the left-hand integrand in (96); by (98), (99), (100), we have

$$\{d\mathfrak{x}\} = M\{d\mathfrak{y}_1\}\{d\mathfrak{y}_2\}, \qquad \{d\mathfrak{y}_1\} = T^{-1}\{d\mathfrak{q}\}, \qquad VM^2 = TP_0,$$

whence

$$\{d\mathfrak{x}\} = (VTP_0^{-1})^{-\frac{1}{2}}\{d\mathfrak{q}\}\{d\mathfrak{y}_2\}.$$

We perform the integration over $\mathfrak{y}_2$, for fixed $\mathfrak{q}$ and variable r. Since $\mathfrak{P}_0[\mathfrak{y}_2] = p = \mathfrak{T}^{-1}[\mathfrak{q}] - r$ is a positive number, we may apply (97), with $\mathfrak{P}_0$, $\mu - \nu$ instead of $\mathfrak{P}$, μ. By (99), we have $p = WT^{-1}$, and the assertion (96) follows readily, for $\lambda = 1$ and arbitrary $\nu < \mu$.

In the remaining case $\lambda > 1$, we apply induction with respect to λ. We split $\mathfrak{X} = (\mathfrak{X}_0, \mathfrak{x})$ and set

$$(\mathfrak{F}, \mathfrak{X}_0) = \mathfrak{F}_0, \qquad \mathfrak{B}[\mathfrak{F}_0] = \mathfrak{W}_0 = \begin{pmatrix} \mathfrak{T} & \mathfrak{Q}_0 \\ \mathfrak{Q}_0' & \mathfrak{R}_0 \end{pmatrix}, \qquad \mathfrak{B}[\mathfrak{F}_0, \mathfrak{x}] = \begin{pmatrix} \mathfrak{W}_0 & \mathfrak{q} \\ \mathfrak{q}' & r \end{pmatrix}.$$

following from Lemmata 5, 17, 20, for all points $\mathfrak{X}$ in $E^*(\mathfrak{F}, \mathfrak{R}_0)$. If we replace the factor α_2 on the right-hand side in (76) by $a_{24} \prod_{k=1}^{m} (1 + h_k^{-n/2})$, then we get the right-hand side in (93); we remark that the latter factor does not depend on $\mathfrak{Z}$. Otherwise, the proof remains the same, and we infer that

$$\Delta_h^* < a_{25}\, \epsilon^{\frac{1}{2}h(h+1-m)} \prod_{k=1}^{m} (1 + h_k^{-n/2}). \tag{94}$$

On the other hand, $h(h + 1 - m) \geqq n(n + 1 - m)$, for $h = 1, \cdots, n$, and

$$T(4\epsilon) \leqq \int_E \operatorname{abs} f(\mathfrak{Z})\, dv_n\,, \tag{95}$$

by (80). Because of $a_{17} = \frac{1}{2}\sqrt{3} > \frac{1}{4}$, the assertion follows from (79), (92), (94), (95), if we replace ϵ by $\dfrac{\epsilon}{4}$.

10. Integration over $\mathfrak{H}$

If $\mathfrak{X} = (x_{kl})$ is a real matrix with independently variable elements, we define $\{d\mathfrak{X}\} = \prod_{k,l} dx_{kl}$; analogously, for symmetric $\mathfrak{X}$, we define $\{d\mathfrak{X}\} = \prod_{k \leqq l} dx_{kl}$. For any complex square matrix designated by a capital Gothic letter, we denote the absolute value of the determinant by the corresponding capital Roman letter; e.g., abs $\mathfrak{T} = T$.

Let $\mathfrak{V}^{(\mu)}$ be a non-singular real symmetric matrix of signature α, $\mu - \alpha$, and let $\mathfrak{F}^{(\mu,\nu)}$ be a real matrix such that $\mathfrak{V}[\mathfrak{F}] = \mathfrak{T}$ has the signature α, $\nu - \alpha$. If $\mathfrak{X}^{(\mu,\lambda)}$ is a variable real matrix, then the formula

$$\mathfrak{V}[\mathfrak{F}, \mathfrak{X}] = \mathfrak{W} = \begin{pmatrix} \mathfrak{T} & \mathfrak{Q} \\ \mathfrak{Q}' & \mathfrak{R} \end{pmatrix}$$

with $\mathfrak{Q} = \mathfrak{F}'\mathfrak{V}\mathfrak{X}$, $\mathfrak{R} = \mathfrak{V}[\mathfrak{X}]$, defines a mapping of the $\mathfrak{X}$-space into the $\mathfrak{Q}$, $\mathfrak{R}$-space. Suppose that $\mathfrak{W}$ is non-singular, then $\nu + \lambda \leqq \mu$, and the signature of $\mathfrak{W}$ is α, $\nu + \lambda - \alpha$.

Lemma 26. *Let D be any domain in the $\mathfrak{Q}$, $\mathfrak{R}$-space such that $\mathfrak{W}$ has the signature α, $\nu + \lambda - \alpha$, let D_0 be the domain in the $\mathfrak{X}$-space which is mapped into D, and define $T = 1$ for $\nu = 0$; then*

$$\int_{D_0} g(\mathfrak{W})\{d\mathfrak{X}\} = \frac{\rho_{\mu-\nu}}{\rho_{\mu-\nu-\lambda}} (VT)^{-\lambda/2} \int_D g(\mathfrak{W})(TW^{-1})^{\frac{1}{2}(\lambda+\nu-\mu+1)}\{d\mathfrak{Q}\}\{d\mathfrak{R}\}, \tag{96}$$

for any integrable function $g(\mathfrak{W})$ of the elements of $\mathfrak{W}$.

Proof. In view of the mean value theorem, it suffices to prove the assertion for $g(\mathfrak{W}) = 1$. We consider first the particular case $\lambda = 1$, $\nu = 0$; then $\alpha = 0$,

– $\mathfrak{B}$ is a positive symmetric matrix $\mathfrak{P}^{(\mu)}$ and $\mathfrak{W} = \mathfrak{R}$ is a negative number $r = -p$. The volume of the ellipsoid $\mathfrak{P}[\mathfrak{x}] \leqq p$ is

$$s(p) = \frac{\pi^{\mu/2}}{\Gamma\left(\dfrac{\mu}{2} + 1\right)} \, |\mathfrak{P}|^{-\frac{1}{2}} p^{\mu/2};$$

moreover, $\pi^{\mu/2}/\Gamma\left(\dfrac{\mu}{2}\right) = \rho_\mu/\rho_{\mu-1}$, by (91); hence

$$\text{(97)} \qquad \int_{\mathfrak{P}[\mathfrak{x}] \leqq a} \{d\mathfrak{x}\} = \int_0^a \frac{ds(p)}{dp}\, dp = \frac{\rho_\mu}{\rho_{\mu-1}} \, |\mathfrak{P}|^{-\frac{1}{2}} \int_0^a p^{\frac{1}{2}\mu - 1}\, dp.$$

This proves (96) for $\lambda = 1$, $\nu = 0$.

Suppose next that $\lambda = 1$, $\nu > 0$; then $\mathfrak{Q} = \mathfrak{q}$ is a column of ν elements and $\mathfrak{R} = r$ is again a number. We choose a real matrix $\mathfrak{F}_1^{(\mu, \mu-\nu)}$ such that $(\mathfrak{F}, \mathfrak{F}_1) = \mathfrak{M}$ is non-singular and set

$$\text{(98)} \qquad \mathfrak{M}^{-1}(\mathfrak{F}, \mathfrak{x}) = \begin{pmatrix} \mathfrak{E}^{(\nu)} & \mathfrak{y}_1 \\ 0 & \mathfrak{y}_2 \end{pmatrix}, \qquad \mathfrak{B}[\mathfrak{M}] = \begin{pmatrix} \mathfrak{T} & \mathfrak{Q}_0 \\ \mathfrak{Q}_0' & \mathfrak{R}_0 \end{pmatrix} = \mathfrak{W}_0,$$

$$\mathfrak{R}_0 - \mathfrak{T}^{-1}[\mathfrak{Q}_0] = -\mathfrak{P}_0, \qquad r - \mathfrak{T}^{-1}[\mathfrak{q}] = -p;$$

then

$$\text{(99)} \qquad \mathfrak{W}_0 = \begin{pmatrix} \mathfrak{T} & 0 \\ 0 & -\mathfrak{P}_0 \end{pmatrix} \begin{bmatrix} \mathfrak{E} & \mathfrak{T}^{-1}\mathfrak{Q}_0 \\ 0 & \mathfrak{E} \end{bmatrix}, \qquad \mathfrak{W} = \begin{pmatrix} \mathfrak{T} & 0 \\ 0 & -p \end{pmatrix} \begin{bmatrix} \mathfrak{E} & \mathfrak{T}^{-1}\mathfrak{q} \\ 0 & 1 \end{bmatrix},$$

$$\text{(100)} \qquad \mathfrak{T}^{-1}\mathfrak{q} = \mathfrak{y}_1 + \mathfrak{T}^{-1}\mathfrak{Q}_0\mathfrak{y}_2, \qquad \mathfrak{P}_0[\mathfrak{y}_2] = p.$$

Since $\mathfrak{B}$ and $\mathfrak{T}$ have the signatures α, $\mu - \alpha$ and α, $\nu - \alpha$, it follows that $\mathfrak{P}_0$ and p are positive. Instead of $\mathfrak{x}$, we introduce the variables $\mathfrak{q}$, $\mathfrak{y}_2$ into the left-hand integrand in (96); by (98), (99), (100), we have

$$\{d\mathfrak{x}\} = M\{d\mathfrak{y}_1\}\{d\mathfrak{y}_2\}, \qquad \{d\mathfrak{y}_1\} = T^{-1}\{d\mathfrak{q}\}, \qquad VM^2 = TP_0,$$

whence

$$\{d\mathfrak{x}\} = (VTP_0^{-1})^{-\frac{1}{2}}\{d\mathfrak{q}\}\{d\mathfrak{y}_2\}.$$

We perform the integration over $\mathfrak{y}_2$, for fixed $\mathfrak{q}$ and variable r. Since $\mathfrak{P}_0[\mathfrak{y}_2] = p = \mathfrak{T}^{-1}[\mathfrak{q}] - r$ is a positive number, we may apply (97), with $\mathfrak{P}_0$, $\mu - \nu$ instead of $\mathfrak{P}$, μ. By (99), we have $p = WT^{-1}$, and the assertion (96) follows readily, for $\lambda = 1$ and arbitrary $\nu < \mu$.

In the remaining case $\lambda > 1$, we apply induction with respect to λ. We split $\mathfrak{X} = (\mathfrak{X}_0, \mathfrak{x})$ and set

$$(\mathfrak{F}, \mathfrak{X}_0) = \mathfrak{F}_0, \qquad \mathfrak{B}[\mathfrak{F}_0] = \mathfrak{W}_0 = \begin{pmatrix} \mathfrak{T} & \mathfrak{Q}_0 \\ \mathfrak{Q}_0' & \mathfrak{R}_0 \end{pmatrix}, \qquad \mathfrak{B}[\mathfrak{F}_0, \mathfrak{x}] = \begin{pmatrix} \mathfrak{W}_0 & \mathfrak{q} \\ \mathfrak{q}' & r \end{pmatrix}.$$

Use (96) with $\lambda = 1$ and $\mathfrak{F}_0$, $\mathfrak{x}$ instead of $\mathfrak{F}$, $\mathfrak{X}$; then

$$(101)\quad \int \{d\mathfrak{X}\} = \frac{\rho_{\mu-\nu-\lambda+1}}{\rho_{\mu-\nu-\lambda}} \int \left((VW_0)^{-\frac{1}{2}} \int (W_0 W^{-1})^{\frac{1}{2}(\lambda+\nu-\mu+1)} \{d\mathfrak{q}\} \{dr\} \right) \{d\mathfrak{X}_0\}.$$

Using (96) once more with $\lambda - 1$, $\mathfrak{F}$, $\mathfrak{X}_0$, $\mathfrak{W}_0$ instead of λ, $\mathfrak{F}$, $\mathfrak{X}$, $\mathfrak{W}$, we obtain

$$(102)\quad \int g(\mathfrak{W}_0)\{d\mathfrak{X}_0\} = \frac{\rho_{\mu-\nu}}{\rho_{\mu-\nu-\lambda+1}} (VT)^{-\frac{1}{2}(\lambda-1)} \int g(\mathfrak{W}_0)(TW_0^{-1})^{\frac{1}{2}(\lambda+\nu-\mu)} \{d\mathfrak{Q}_0\} \{d\mathfrak{R}_0\}.$$

Since $W = W_0(\mathfrak{W}_0^{-1}[\mathfrak{q}] - r)$, the outer right-hand integrand in (101) is a function $g(\mathfrak{W}_0)$, and the assertion (96) follows from (101), (102), for arbitrary λ. The lemma is now completely proved.

In the specical case $\mathfrak{W} = -\mathfrak{E}$, $\nu = 0$, $\lambda = \mu$, $g(\mathfrak{W}) = 1$, the formula (96) becomes

$$\int_{D_0} \{d\mathfrak{X}\} = \rho_\mu \int_D W^{-\frac{1}{2}} \{d\mathfrak{W}\},$$

with $[\mathfrak{X}] = \mathfrak{W}$; this proves that ρ_μ is the volume of the space of the μ-rowed orthogonal matrices, computed with the volume element (1).

Both sides in the assertion of Theorem 1 are class-invariants, i.e., invariants with respect to the transformation $\mathfrak{S} \to \mathfrak{S}[\mathfrak{U}]$ for unimodular $\mathfrak{U}$; therefore it suffices to prove Theorem 1 for $\mathfrak{S}[\mathfrak{U}]$ instead of $\mathfrak{S}$, with suitably chosen $\mathfrak{U}$.

LEMMA 27. *Let $\mathfrak{S}$ have the signature r, $m - r$; then there exists a primitive $\mathfrak{F}^{(m,m-r)}$ such that* $-\mathfrak{S}[\mathfrak{F}] > 0$.

PROOF. Choose the real $\mathfrak{K}^{(m)}$ such that $\mathfrak{S}[\mathfrak{K}] = [p_1, \cdots, p_m]$, $p_k = -1$ $(k = 1, \cdots, m - r)$, $p_k = 1$ $(k = m - r + 1, \cdots, m)$, and let $\mathfrak{F}_0$ denote the matrix of the first $m - r$ columns in $\mathfrak{K}$; then $-\mathfrak{S}[\mathfrak{F}_0] = \mathfrak{E}^{(m-r)} > 0$. Consequently, there exists also a rational $\mathfrak{F}_1^{(m,m-r)}$ with $-\mathfrak{S}[\mathfrak{F}_1] > 0$. Determine a non-singular rational $\mathfrak{Q}^{(m-r)}$ such that $\mathfrak{F}_1\mathfrak{Q} = \mathfrak{F}$ is primitive; then $\mathfrak{F}$ has the required property.

Since any primitive matrix can be completed to a unimodular matrix, it follows from Lemma 27 that the condition (14) of Lemma 4 may be satisfied within the class of $\mathfrak{S}$. Henceforth we shall suppose that $\mathfrak{S}$ already fulfills (14). By Lemma 4, we have then the parametric representation (15) of the space H of all positive symmetric $\mathfrak{H}$ satisfying $\mathfrak{H}\mathfrak{S}^{-1}\mathfrak{H} = \mathfrak{S}$, namely

$$(103)\quad \mathfrak{H} = 2\mathfrak{W}^{-1}[\mathfrak{X}'\mathfrak{S}] - \mathfrak{S}, \qquad \mathfrak{S}[\mathfrak{X}] = \mathfrak{W} > 0, \qquad \mathfrak{X} = \begin{pmatrix} \mathfrak{Y} \\ \mathfrak{E} \end{pmatrix},$$

with variable real $\mathfrak{Y}^{(m-r,r)}$. We introduce the volume element

$$(104)\quad dv_H = S^{r/2} W^{-m/2} \{d\mathfrak{Y}\}.$$

Let $\mathfrak{X}_0^{(r)}$ and $\mathfrak{X}_2^{(m,m-r)}$ be variable real matrices and define

$$(105)\quad \mathfrak{X}_1 = (\mathfrak{X}\mathfrak{X}_0, \mathfrak{X}_2), \qquad \mathfrak{S}[\mathfrak{X}_1] = \mathfrak{W}_1 = \begin{pmatrix} \mathfrak{W}_0 & \mathfrak{Q} \\ \mathfrak{Q}' & \mathfrak{R} \end{pmatrix}, \qquad \mathfrak{W}_0 = \mathfrak{S}[\mathfrak{X}\mathfrak{X}_0] = \mathfrak{W}[\mathfrak{X}_0].$$

Choose a domain D in the $\mathfrak{W}_1$-space such that $\mathfrak{W}_0 > 0$ and that $\mathfrak{W}_1$ has the signature $r, m - r$; also, let H_0 be a domain in H. In view of (103), (105), the $\mathfrak{X}_1$-space is mapped into the $\mathfrak{H}$, $\mathfrak{W}_1$-space; let D_0 be the domain of all $\mathfrak{X}_1$ such that $\mathfrak{H}$ lies in H_0 and $\mathfrak{W}_1$ in D.

LEMMA 28.

$$\int_{D_0} \{d\mathfrak{X}_1\} = \rho_r \rho_{m-r} S^{-m/2} \int_{H_0} dv_H \int W_1^{-\frac{1}{2}} \{d\mathfrak{W}_1\}.$$

PROOF. Suppose first that $\mathfrak{X}$ and $\mathfrak{X}_0$ are given; then also $\mathfrak{W}_0$ is fixed. Let $D(\mathfrak{W}_0)$ be the corresponding cross-section of D, and let $D_0(\mathfrak{W}_0)$ be the set of $\mathfrak{X}_2$ mapped into $D(\mathfrak{W}_0)$. Since $\mathfrak{W}_0$ has the signature r, 0, we may apply Lemma 26 with $\mu = m$, $\nu = r$, $\lambda = m - r$ and $\mathfrak{X}_2$, $\mathfrak{S}$, $\mathfrak{W}_0$, $\mathfrak{W}_1$ instead of $\mathfrak{X}$, $\mathfrak{B}$, $\mathfrak{T}$, $\mathfrak{W}$; hence

$$\int_{D_0(\mathfrak{W}_0)} \{d\mathfrak{X}_2\} = \rho_{m-r}(SW_0)^{\frac{1}{2}(r-m)} \int_{D(\mathfrak{W}_0)} (W_0 W_1^{-1})^{\frac{1}{2}} \{d\mathfrak{Q}\}\{d\mathfrak{R}\} = f(\mathfrak{W}_0), \tag{106}$$

say. For variable $\mathfrak{X}$, $\mathfrak{X}_0$, we have

$$\{d\mathfrak{X}_1\} = X_0^{m-r}\{d\mathfrak{Y}\}\{d\mathfrak{X}_0\}, \qquad X_0 = (W_0 W^{-1})^{\frac{1}{2}}, \tag{107}$$

by (105). Using Lemma 26 once more, with $\mu = r$, $\nu = 0$, $\lambda = r$ and $\mathfrak{X}_0$, $\mathfrak{W}$, $\mathfrak{W}_0$, $X_0^{m-r} f(\mathfrak{W}_0)$ instead of $\mathfrak{X}$, $\mathfrak{B}$, $\mathfrak{W}$, $g(\mathfrak{W})$, we obtain

$$\int f(\mathfrak{W}_0) X_0^{m-r}\{d\mathfrak{X}_0\} = \rho_r W^{-m/2} \int f(\mathfrak{W}_0) W_0^{\frac{1}{2}(m-r-1)}\{d\mathfrak{W}_0\}, \tag{108}$$

both integrations extended over the whole space. Integrating (106) over $\mathfrak{X}_1$ and using (106), (107), (108), we get the desired result.

Let $\Omega(\mathfrak{S})$ be the group of all real $\mathfrak{U}$ with $\mathfrak{S}[\mathfrak{U}] = \mathfrak{S}$. The transformation $\mathfrak{H} \to \mathfrak{H}[\mathfrak{U}]$ maps H onto itself, for all $\mathfrak{U}$ in $\Omega(\mathfrak{S})$; because of $\{d(\mathfrak{U}\mathfrak{X}_1)\} = \{d\mathfrak{X}_1\}$, we infer from Lemma 28 that the volume element dv_H is invariant under this mapping. It can be proved that dv_H is the volume element for the invariant metric defined by $ds^2 = \frac{1}{8}\sigma(\mathfrak{H}^{-1}d\mathfrak{H}\mathfrak{H}^{-1}d\mathfrak{H})$; however, we do not need this relation, and we omit the proof.

On the other hand, the definition of the volume element dv in (1) implies the formula

$$\int_{D_0} \{d\mathfrak{X}_1\} = S^{\frac{1}{2}} \int_D W_1^{-\frac{1}{2}} v(D_0, \mathfrak{W}_1)\{d\mathfrak{W}_1\},$$

where $v(D_0, \mathfrak{W}_1)$, for any fixed $\mathfrak{W}_1$, is the volume of the cross-section $\mathfrak{S}[\mathfrak{X}_1] = \mathfrak{W}_1$ of D_0, computed with the volume element (1). It follows from Lemma 28 that

$$v(D_0, \mathfrak{W}_1) = \rho_r \rho_{m-r} S^{-\frac{1}{2}(m+1)} \int_{H_0} dv_H. \tag{109}$$

Consider now the group $\Gamma(\mathfrak{S})$ of all units of $\mathfrak{S}$, i.e., the subgroup of all integral $\mathfrak{U}$ in $\Omega(\mathfrak{S})$. Plainly, $\mathfrak{U}$ and $-\mathfrak{U}$ lead to the same mapping $\mathfrak{H} \to \mathfrak{H}[\mathfrak{U}]$ on H.

Identifying $\mathfrak{U}$ and $-\mathfrak{U}$, we obtain a factor group $\Gamma^*(\mathfrak{S})$ of $\Gamma(\mathfrak{S})$. We have to use some known properties of $\Gamma^*(\mathfrak{S})$; the proofs are contained in my paper: *Einheiten quadratischer Formen*, Abh. Math. Sem. Hansischen Univ. 13, pp. 209–239 (1940).

The group $\Gamma^*(\mathfrak{S})$ is discontinuous on H and possesses there a fundamental domain H_0 of finite volume $v_H(\mathfrak{S})$, measured with the volume element (104); the trivial case of a decomposable binary quadratic form being excepted. The image F_0 of H_0 on the cross-section $\mathfrak{S}[\mathfrak{X}_1] = \mathfrak{W}_1$ of D_0, for fixed $\mathfrak{W}_1$, admits the mapping $\mathfrak{X}_1 \to -\mathfrak{X}_1$ onto itself; consequently, F_0 is the double of a fundamental domain for the representation $\mathfrak{X}_1 \to \mathfrak{U}\mathfrak{X}_1$ of $\Gamma(\mathfrak{S})$. In view of the definition of $\rho(\mathfrak{S})$ in the introduction, (109) leads to the formula

$$\rho(\mathfrak{S}) = \tfrac{1}{2}\rho_r\rho_{m-r}S^{-\frac{1}{2}(m+1)}v_H(\mathfrak{S}). \tag{110}$$

Lemma 29. *Let $m_k(\mathfrak{H}) = h_k$ $(k = 1, \cdots, m)$ and suppose that $2n + 2 < m$; then*

$$\int_{H_0}\prod_{k=1}^{m}(1 + h_k^{-n/2})\,dv_H < a_{26}.$$

Proof. Obviously, $\Gamma(\mathfrak{S})$ and H are not changed if $\mathfrak{S}$ is replaced by $-\mathfrak{S}$; therefore we may suppose that $r \leqq m - r$.

Let l be an integer in the interval $0 \leqq l \leqq r$, and let $g_1, \cdots, g_l$ denote arbitrary positive integers. Let $H(l; g_1, \cdots, g_l)$ be the set of all $\mathfrak{H}$ in the fundamental domain H_0 satisfying the following conditions:

$$-g_l < \log h_l \leqq 1 - g_l, \qquad -g_k < \log\frac{h_k}{h_{k+1}} \leqq 1 - g_k \quad (k = 1, \cdots, l-1),$$

$$h_l h_{m-l} < 1, \qquad h_k h_{m-k} \geqq 1 \quad (l < k < m - l);$$

in case $l = 0$, these conditions mean $h_k h_{m-k} \geqq 1$ $(k = 1, \cdots, m-1)$. Plainly,

$$\prod_{k=1}^{l} h_k \geqq e^{-\sum_{k=1}^{l} k g_k}; \tag{111}$$

on the other hand, it is known that

$$h_k > a_{27} \qquad (k = l+1, \cdots, m), \tag{112}$$

that the sets $H(l; g_1, \cdots, g_l)$, for $g_1, \cdots, g_l = 1, 2, \cdots$ and $l = 0, \cdots, r$, cover H_0 completely, and that their volumes satisfy the inequality

$$\int_{H(l;g_1,\cdots,g_l)} dv_H < a_{28}\, e^{-\frac{1}{2}\sum_{k=1}^{l} k(m-k-1)g_k}. \tag{113}$$

It follows from (111), (112), (113) that

$$\int_{H_0}\prod_{k=1}^{m}(1 + h_k^{-n/2})\,dv_H < a_{29}\sum_{l=0}^{r}\sum_{g_1,\cdots,g_l} e^{-\frac{1}{2}\sum_{k=1}^{l} k(m-n-k-1)g_k}, \tag{114}$$

where the inner sum means 1 for $l = 0$ and $g_1, \cdots, g_l$ run independently over all positive integers. Since $2n + 2 < m$ and $r \leqq m - r$, we have

$$m - n - k - 1 \geqq m - n - l - 1 \geqq m - n - r - 1 > m - \frac{m}{2} - \frac{m}{2} = 0;$$

hence the sum in (114) converges, and the assertion is proved.

We multiply the formula of Theorem 4 by dv_H and integrate over the fundamental domain H_0. In view of Theorem 5 and Lemma 29, the integration over $\mathfrak{H}$ and the passage to the limit $\epsilon \to 0$ may be interchanged. By (110), we obtain

LEMMA 30. *Let $2n + 2 < m$, then*

$$\lim_{\epsilon \to 0} \epsilon^{\frac{1}{2}n(m-n-1)} \sum_{\mathfrak{S}[\mathfrak{G}] = \mathfrak{T}} \int_{H_0} e^{-\frac{1}{2}\pi\epsilon\sigma(\mathfrak{H}[\mathfrak{G}])}\, dv_H$$
$$= \frac{2\rho_{m-2n-1}}{\rho_{r-n}\,\rho_{m-r-n}\,\rho_{m-n-1}}\, S^{\frac{1}{2}(m-n+1)}\, \rho(\mathfrak{S}) \prod_p \alpha_p(\mathfrak{S}, \mathfrak{T}).$$

If $\mathfrak{T} = 0$, then $\mathfrak{G} = 0$ is a solution of $\mathfrak{S}[\mathfrak{G}] = \mathfrak{T}$. Obviously, Lemma 30 remains true if we cancel this term; henceforth we shall suppose that $\mathfrak{G} \neq 0$.

Consider now the subgroup $\Gamma(\mathfrak{S}, \mathfrak{G})$ of $\Gamma(\mathfrak{S})$ consisting of all $\mathfrak{U}$ in $\Gamma(\mathfrak{S})$ satisfying $\mathfrak{U}\mathfrak{G} = \mathfrak{G}$. Since $-\mathfrak{E}$ does not belong to $\Gamma(\mathfrak{S}, \mathfrak{G})$, we may choose a sequence $\mathfrak{U}_1, \mathfrak{U}_2, \cdots$ of elements in $\Gamma(\mathfrak{S})$ such that $\mathfrak{U}_k, -\mathfrak{U}_k$ $(k = 1, 2, \cdots)$ represent the left cosets of $\Gamma(\mathfrak{S}, \mathfrak{G})$ relative to $\Gamma(\mathfrak{S})$. Let $H(\mathfrak{G})$ denote the union of all images of H_0 under the transformations $\mathfrak{H} \to \mathfrak{H}[\mathfrak{U}_k]$ $(k = 1, 2, \cdots)$; plainly, $H(\mathfrak{G})$ is a fundamental domain for $\Gamma(\mathfrak{S}, \mathfrak{G})$ in H. On the other hand, the set of all $\pm\mathfrak{U}_k\mathfrak{G}$ $(k = 1, 2, \cdots)$ consists of the solutions $\mathfrak{X}$ of $\mathfrak{S}[\mathfrak{X}] = \mathfrak{T}$ which are associate with the given solution $\mathfrak{G}$, relative to $\Gamma(\mathfrak{S})$. Hence

$$\sum_{\mathfrak{G}_1} \int_{H_0} e^{-\frac{1}{2}\pi\epsilon\sigma(\mathfrak{H}[\mathfrak{G}_1])}\, dv_H = 2 \int_{H(\mathfrak{G})} e^{-\frac{1}{2}\pi\epsilon\sigma(\mathfrak{H}[\mathfrak{G}])}\, dv_H, \tag{115}$$

where $\mathfrak{G}_1$ runs over all matrices associate with $\mathfrak{G}$.

Let t be the rank of $\mathfrak{T}^{(n)}$. For the proof of Theorem 1 we may suppose that $\mathfrak{T} = \begin{pmatrix} \mathfrak{T}_0 & 0 \\ 0 & 0 \end{pmatrix}$, with non-singular $\mathfrak{T}_0^{(t)}$. Suppose that h is the rank of a given solution $\mathfrak{G}$ of $\mathfrak{S}[\mathfrak{G}] = \mathfrak{T}$; plainly $t \leqq h \leqq n$.

LEMMA 31. *There exist an integral $\mathfrak{C}^{(m,h)}$ of rank h and a primitive $\mathfrak{F}^{(h-t,n-t)}$ such that*

$$\mathfrak{G} = \mathfrak{C}\mathfrak{P}, \qquad \mathfrak{P} = \begin{pmatrix} \mathfrak{E}^{(t)} & 0 \\ 0 & \mathfrak{F} \end{pmatrix}.$$

PROOF. Any integral $\mathfrak{G}^{(m,n)}$ of rank h can be expressed in the form $\mathfrak{G} = \mathfrak{C}\mathfrak{P}$, with integral $\mathfrak{C}^{(m,h)}$ of rank h and primitive $\mathfrak{P}^{(h,n)}$, and $\mathfrak{P}$ is determined up to a unimodular factor $\mathfrak{V}^{(h)}$ on the left side. Let s be the rank of $\mathfrak{S}[\mathfrak{C}]$; since $\mathfrak{C}$, $\mathfrak{P}$ may be replaced by $\mathfrak{C}\mathfrak{V}^{-1}$, $\mathfrak{V}\mathfrak{P}$, we may suppose that also $\mathfrak{S}[\mathfrak{C}] = \begin{pmatrix} \mathfrak{T}_1 & 0 \\ 0 & 0 \end{pmatrix}$, with non-singular $\mathfrak{T}_1^{(s)}$. Set $\mathfrak{P} = \begin{pmatrix} \mathfrak{P}_1 & \mathfrak{P}_{12} \\ \mathfrak{P}_{21} & \mathfrak{F} \end{pmatrix}$, where $\mathfrak{P}_1$ has s rows and t columns;

then $\mathfrak{T}_1[\mathfrak{P}_1, \mathfrak{P}_{12}] = \mathfrak{T}$. The matrix $(\mathfrak{P}_1, \mathfrak{P}_{12})$ has the rank $s \leqq n$ and n columns; hence $s = t$. Furthermore, $\mathfrak{T}_1[\mathfrak{P}_1] = \mathfrak{T}_0$ and $\mathfrak{P}_1'\mathfrak{T}_1\mathfrak{P}_{12} = 0$, whence $\mathfrak{P}_{12} = 0$. It follows that $\mathfrak{P}_1$ and $\mathfrak{V} = \begin{pmatrix} \mathfrak{P}_1 & 0 \\ \mathfrak{P}_{21} & \mathfrak{E} \end{pmatrix}^{-1}$ are unimodular. Replacing $\mathfrak{C}$, $\mathfrak{P}$ by $\mathfrak{C}\mathfrak{V}^{-1}$, $\mathfrak{V}\mathfrak{P}$, we obtain the assertion.

Obviously, $\Gamma(\mathfrak{S}, \mathfrak{G}) = \Gamma(\mathfrak{S}, \mathfrak{C})$, and we may choose $H(\mathfrak{C}) = H(\mathfrak{G})$. Let $\rho(\mathfrak{S}, \mathfrak{C})$ be the volume of a fundamental domain of $\Gamma(\mathfrak{S}, \mathfrak{C})$ computed with the volume element (1), according to the definition in the introduction.

Lemma 32. *Suppose that $n \leqq r$, $n \leqq m - r$, $2n + 2 < m$, then*

$$\epsilon^{\frac{1}{2}h(m-h-1)} \int_{H(\mathfrak{G})} e^{-\frac{1}{2}\pi\epsilon\sigma(\mathfrak{H}[\mathfrak{G}])}\, dv_H =$$

$$\frac{\rho_{m-2h+t-1}}{\rho_{r-h}\,\rho_{m-r-h}\,\rho_{m-h-1}} S^{\frac{1}{2}(m-h+1)} |[\mathfrak{F}']|^{\frac{1}{2}(h-m+1)} \rho(\mathfrak{S}, \mathfrak{C}) \int_{\mathfrak{Z}>0,\, \mathfrak{Z}+\epsilon\mathfrak{T}_0>0} e^{-\pi\sigma(\mathfrak{Z}+\frac{1}{2}\epsilon\mathfrak{T}_0)} |\mathfrak{Z}|^{\frac{1}{2}(m-r-h-1)}$$

$$|\mathfrak{Z} + \epsilon\mathfrak{T}_0|^{\frac{1}{2}(r-h-1)}\{d\mathfrak{Z}\},$$

where the integration is extended over all real symmetric $\mathfrak{Z}^{(t)}$ satisfying $\mathfrak{Z} > 0$ and $\mathfrak{Z} + \epsilon\mathfrak{T}_0 > 0$.

Proof. Let $\mathfrak{X}_1^{(m,m-h)}$ be variable in a domain D_0, and let D be the image of D_0 in the $\mathfrak{Q}$, $\mathfrak{R}$-space determined by $\mathfrak{S}[\mathfrak{C}, \mathfrak{X}_1] = \mathfrak{W}_1 = \begin{pmatrix} \mathfrak{S}_1 & \mathfrak{Q} \\ \mathfrak{Q}' & \mathfrak{R} \end{pmatrix}$, with $\mathfrak{S}_1 = \mathfrak{S}[\mathfrak{C}] = \begin{pmatrix} \mathfrak{T}_0 & 0 \\ 0 & 0 \end{pmatrix}$. Denote by $V(D_0, \mathfrak{W}_1)$ the volume of the cross-section of D_0, for fixed $\mathfrak{W}_1$, computed with the volume element (1); then

$$\int_{D_0} \{d\mathfrak{X}_1\} = S^{\frac{1}{2}} \int_D W_1^{-\frac{1}{2}} V(D_0, \mathfrak{W}_1)\{d\mathfrak{Q}\}\{d\mathfrak{R}\}. \tag{116}$$

Since $h \leqq n \leqq m - r$, we may split $\mathfrak{X}_1 = (\mathfrak{X}\mathfrak{X}_0, \mathfrak{X}_2)$, $\mathfrak{X} = \begin{pmatrix} \mathfrak{Y} \\ \mathfrak{E} \end{pmatrix}$, where $\mathfrak{X}_2$ has $m - h - r$ columns. Set again $\mathfrak{S}[\mathfrak{X}] = \mathfrak{W}$, $2\mathfrak{W}^{-1}[\mathfrak{X}'\mathfrak{S}] - \mathfrak{S} = \mathfrak{H}$, and define D_0 in the following way: D is a given domain in the space of the matrices $\mathfrak{Q}$, $\mathfrak{R}$ such that $\mathfrak{W}_1 = \begin{pmatrix} \mathfrak{S}_1 & \mathfrak{Q} \\ \mathfrak{Q}' & \mathfrak{R} \end{pmatrix}$ has the signature r, $m - r$, and D_0 consists of all points $\mathfrak{X}_1$ mapped into D and subjected to the condition that $\mathfrak{H}$ lies in $H(\mathfrak{C})$.

For all elements $\mathfrak{U}$ of $\Gamma(\mathfrak{S}, \mathfrak{C})$ and any given $\mathfrak{W}_1$, the transformation $\mathfrak{X}_1 \to \mathfrak{U}\mathfrak{X}_1$ maps the surface $\mathfrak{S}[\mathfrak{C}, \mathfrak{X}_1] = \mathfrak{W}_1$ onto itself; plainly, the intersection $J(\mathfrak{C}, \mathfrak{W}_1)$ of D_0 with this surface is a fundamental domain for $\Gamma(\mathfrak{S}, \mathfrak{C})$. Since $\{d\mathfrak{X}_1\} = \{d(\mathfrak{U}\mathfrak{X}_1)\}$, for any unimodular $\mathfrak{U}^{(m)}$, it follows from (116) that the volume $V(D_0, \mathfrak{W}_1)$ of $J(\mathfrak{C}, \mathfrak{W}_1)$ does not depend on the particular choice of the fundamental domain $H(\mathfrak{C})$. On the other hand, if $\mathfrak{W}_1$ and $\mathfrak{W}_1^* = \begin{pmatrix} \mathfrak{S}_1 & \mathfrak{Q}^* \\ \mathfrak{Q}^{*\prime} & \mathfrak{R}^* \end{pmatrix}$ are any two points in D, then there exists a real matrix $\mathfrak{A}^{(m)} = \begin{pmatrix} \mathfrak{E}^{(h)} & * \\ 0 & * \end{pmatrix}$ satisfying $\mathfrak{W}_1^* = \mathfrak{W}_1[\mathfrak{A}]$,

and the transformation $(\mathfrak{C}, \mathfrak{X}_1) \to (\mathfrak{C}, \mathfrak{X}_1)\mathfrak{A} = (\mathfrak{C}, \mathfrak{X}_1^*)$ maps a fundamental domain on $\mathfrak{S}[\mathfrak{C}, \mathfrak{X}_1] = \mathfrak{W}_1$ onto a fundamental domain on $\mathfrak{S}[\mathfrak{C}, \mathfrak{X}_1^*] = \mathfrak{W}_1^*$. Since

$$\{d\mathfrak{X}_1^*\} = A^m\{d\mathfrak{X}_1\}, \quad \{d\mathfrak{Q}^*\} = A^h\{d\mathfrak{Q}\}, \quad \{d\mathfrak{R}^*\} = A^{m-h+1}\{d\mathfrak{R}\}, \quad W_1^* = A^2W_1,$$

we infer from (116) that $V(D_0, \mathfrak{W}_1) = \rho(\mathfrak{S}, \mathfrak{C})$ is also independent of $\mathfrak{W}_1$.

Define $\mathfrak{S}[\mathfrak{C}, \mathfrak{X}\mathfrak{X}_0] = \begin{pmatrix} \mathfrak{S}_1 & \mathfrak{L}_0 \\ \mathfrak{L}_0' & \mathfrak{W}_0 \end{pmatrix} = \mathfrak{W}_2$; let D^* be a domain in the $\mathfrak{L}_0$, $\mathfrak{W}_0$-space such that $\mathfrak{W}_2$ has the signature r, h, and let D_0^* be the corresponding $\mathfrak{X}\mathfrak{X}_0$-domain, $\mathfrak{H}$ lying in $H(\mathfrak{C})$. Applying Lemma 26 with $\mu = m$, $\nu = h + r$, $\lambda = m - h - r$ and $\mathfrak{X}_2$, $\mathfrak{S}$, $\mathfrak{W}_2$, $\mathfrak{W}_1$ instead of $\mathfrak{X}$, $\mathfrak{B}$, $\mathfrak{T}$, $\mathfrak{W}$, we obtain from (116) the formula

$$\int_{D_0^*} g(\mathfrak{W}_2)\{d(\mathfrak{X}\mathfrak{X}_0)\} = \frac{\rho(\mathfrak{S}, \mathfrak{C})}{\rho_{m-r-h}} S^{\frac{1}{2}(m-h-r+1)} \int_{D^*} g(\mathfrak{W}_2) W_2^{\frac{1}{2}(m-h-r-1)}\{d\mathfrak{L}_0\}\{d\mathfrak{W}_0\},$$

for any integrable function $g(\mathfrak{W}_2)$ of the elements of $\mathfrak{W}_2$.

In particular we choose D^* in the following way: $\mathfrak{W}_0$ lies in a given domain Δ of r-rowed positive symmetric matrices and $\mathfrak{L}_0^{(h,r)}$ runs over all real matrices such that $\mathfrak{W}_0^{-1}[\mathfrak{L}_0'] - \mathfrak{S}_1 = \mathfrak{Z} > 0$. Then D_0^* consists of all $\mathfrak{X}\mathfrak{X}_0$ such that $\mathfrak{H}$ lies in $H(\mathfrak{C})$ and $\mathfrak{W}[\mathfrak{X}_0] = \mathfrak{W}_0$ in Δ, and we have

$$\{d(\mathfrak{X}\mathfrak{X}_0)\} = (W_0W^{-1})^{\frac{1}{2}(m-r)}\{d\mathfrak{Y}\}\{d\mathfrak{X}_0\}, \qquad W_2 = W_0Z.$$

Set $\mathfrak{L}_0 = \mathfrak{L}\mathfrak{X}_0$, then

$$\mathfrak{Z} = \mathfrak{W}^{-1}[\mathfrak{L}'] - \mathfrak{S}_1 = \tfrac{1}{2}(\mathfrak{H} - \mathfrak{S})[\mathfrak{C}], \qquad \{d\mathfrak{L}_0\} = (W_0W^{-1})^{h/2}\{d\mathfrak{L}\}.$$

Apply Lemma 26 once more with $\mu = r$, $\nu = 0$, $\lambda = r$ and $\mathfrak{X}_0$, $\mathfrak{W}$, $\mathfrak{W}_0$ instead of $\mathfrak{X}$, $\mathfrak{B}$, $\mathfrak{W}$; in view of (104), we obtain

$$\int_{H(\mathfrak{C})} g(\mathfrak{Z})\, dv_H = \frac{\rho(\mathfrak{S}, \mathfrak{C})}{\rho_r\,\rho_{m-r-h}} S^{\frac{1}{2}(m-h+1)} \int_{\mathfrak{Z}>0} g(\mathfrak{Z}) W^{-h/2} Z^{\frac{1}{2}(m-h-r-1)}\{d\mathfrak{L}\},$$

for any integrable function $g(\mathfrak{Z})$ of $\mathfrak{Z}$.

Since $h \leqq n \leqq r$, we may use Lemma 21 a third time with $\mu = r$, $\nu = 0$, $\lambda = h$ and $\mathfrak{L}'$, $\mathfrak{W}^{-1}$, $\mathfrak{Z} + \mathfrak{S}_1$ instead of $\mathfrak{X}$, $\mathfrak{B}$, $\mathfrak{W}$; replacing $\mathfrak{Z}$ by $\epsilon^{-1}\mathfrak{Z}$ on the right-hand side, we infer that

$$\epsilon^{\frac{1}{2}h(m-h-1)} \int_{H(\mathfrak{C})} g(\mathfrak{Z})\, dv_H = \frac{\rho(\mathfrak{S}, \mathfrak{C})}{\rho_{r-h}\,\rho_{m-r-h}} S^{\frac{1}{2}(m-h+1)} \tag{117}$$
$$\int_{\mathfrak{Z}>0,\, \mathfrak{Z}+\epsilon\mathfrak{S}_1>0} g(\epsilon^{-1}\mathfrak{Z}) Z^{\frac{1}{2}(m-h-r-1)} \,|\, \mathfrak{Z} + \epsilon\mathfrak{S}_1 \,|^{\frac{1}{2}(r-h-1)}\{d\mathfrak{Z}\}.$$

By Lemma 31, we have $\mathfrak{H}[\mathfrak{G}] = 2\mathfrak{Z}[\mathfrak{P}] + \mathfrak{T}$. We choose now

$$g(\mathfrak{Z}) = e^{-\frac{1}{2}\pi\epsilon\sigma(\mathfrak{H}[\mathfrak{G}])} = e^{-\pi\epsilon\sigma(\mathfrak{Z}[\mathfrak{P}]+\frac{1}{2}\mathfrak{T})} \tag{118}$$

and set $\mathfrak{Z} = \begin{pmatrix} \mathfrak{Z}_1 & 0 \\ 0 & \mathfrak{Z}_2 \end{pmatrix}\begin{bmatrix} \mathfrak{E} & 0 \\ \mathfrak{X}_3 & \mathfrak{E} \end{bmatrix}$, with variable $\mathfrak{Z}_1^{(t)}$, $\mathfrak{Z}_2^{(h-t)}$, $\mathfrak{X}_3^{(h-t,t)}$; then

$$\{d\mathfrak{Z}\} = Z_2^t\{d\mathfrak{Z}_1\}\{d\mathfrak{Z}_2\}\{d\mathfrak{X}_3\}, \qquad Z = Z_1Z_2, \tag{119}$$

$$|\, \mathfrak{Z} + \epsilon\mathfrak{S}_1 \,| = Z_2 \,|\, \mathfrak{Z}_1 + \mathfrak{T}_0 \,|, \quad \sigma(\mathfrak{Z}[\mathfrak{P}]) = \sigma(\mathfrak{Z}_1) + \sigma(\mathfrak{Z}_2[\mathfrak{X}_3]) + \sigma(\mathfrak{Z}_2[\mathfrak{F}]). \tag{120}$$

The conditions $\mathfrak{Z} > 0$, $\mathfrak{Z} + \epsilon\mathfrak{S}_1 > 0$ are satisfied if and only if $\mathfrak{Z}_2 > 0$, $\mathfrak{Z}_1 > 0$, $\mathfrak{Z}_1 + \epsilon\mathfrak{T}_0 > 0$; consequently, $\mathfrak{X}_3$ runs over all real matrices. Since

$$\int_{-\infty}^{+\infty} e^{-\pi\sigma(\mathfrak{Z}_2[\mathfrak{X}_3])}\{d\mathfrak{X}_3\} = Z_2^{-t/2}$$

and

$$\int_{\mathfrak{Z}_2>0} e^{-\pi\sigma([\mathfrak{F}']\mathfrak{Z}_2)}\, Z_2^{\frac{1}{2}(m+t)-h-1}\{d\mathfrak{Z}_2\} = \frac{\rho_{m-2h+t-1}}{\rho_{m-h-1}}\,|\,[\mathfrak{F}']\,|^{\frac{1}{2}(h-m+1)},$$

the assertion follows from (117), (118), (119), (120).

By (115) and Lemma 32, we have

$$\epsilon^{\frac{1}{2}n(m-n-1)} \sum_{\mathfrak{S}[\mathfrak{G}]=\mathfrak{T}} \int_{H_0} e^{-\frac{1}{2}\pi\epsilon\sigma(\mathfrak{H}[\mathfrak{G}])} = \sum_{h=t}^{n} f_h\,\varphi_h(\epsilon), \tag{121}$$

where

$$\varphi_h(\epsilon) = \epsilon^{\frac{1}{2}n(m-n-1)-\frac{1}{2}h(m-h-1)}$$

$$\cdot\int_{\mathfrak{Z}>0,\,\mathfrak{Z}+\epsilon\mathfrak{T}_0>0} e^{-\pi\sigma(\mathfrak{Z}+\frac{1}{2}\epsilon\mathfrak{T}_0)}\,|\,\mathfrak{Z}\,|^{\frac{1}{2}(m-r-h-1)}\,|\mathfrak{Z}+\epsilon\mathfrak{T}_0\,|^{\frac{1}{2}(r-h-1)}\,\{d\mathfrak{Z}\} \tag{122}$$

and f_h is a non-negative quantity independent of ϵ; in particular,

$$f_n = \frac{2\rho_{m-2n+t-1}}{\rho_{r-n}\,\rho_{m-r-n}\,\rho_{m-n-1}}\, S^{\frac{1}{2}(m-n+1)} \sum_{\mathfrak{G}} \rho(\mathfrak{S}, \mathfrak{G}), \tag{123}$$

where $\mathfrak{G}$ runs over a complete set of non-associate $\mathfrak{G}$ of rank n, with $\mathfrak{S}[\mathfrak{G}] = \mathfrak{T}$. The finiteness of f_n is implied in Lemma 30; in particular, all $\rho(\mathfrak{S}, \mathfrak{C})$ are finite.

Suppose that $r \geqq m - r$, then $r - h - 1 \geqq r - n - 1 > 0$, because of $n + 1 < \dfrac{m}{2}$. It follows from the inequality

$$|\,\mathfrak{Z} + \epsilon\mathfrak{T}_0\,|^{\frac{1}{2}(r-h-1)} < a_{30}\, e^{\sigma(\mathfrak{Z}+\epsilon\mathfrak{T}_0)}$$

that the integral in (122) converges uniformly for $0 < \epsilon < 1$. If $r < m - r$, then we replace $\mathfrak{Z} + \epsilon\mathfrak{T}_0$, $\mathfrak{Z}$ by $\mathfrak{Z}$, $\mathfrak{Z} - \epsilon\mathfrak{T}_0$ and use the inequality

$$|\,\mathfrak{Z} - \epsilon\mathfrak{T}_0\,|^{\frac{1}{2}(m-r-h-1)} < a_{31}\, e^{\sigma(\mathfrak{Z}-\epsilon\mathfrak{T}_0)},$$

arriving at the same conclusion.

Since $\dfrac{h}{2}(m - h - 1) < \dfrac{n}{2}(m - n - 1)$, for $h = 0, \cdots, n - 1$, we infer that $\lim\limits_{\epsilon\to 0} \varphi_h(\epsilon) = 0$ for $h < n$. Moreover

$$\lim_{\epsilon\to 0} \varphi_n(\epsilon) = \varphi_n(0) = \int_{\mathfrak{Z}>0} e^{-\pi\sigma(\mathfrak{Z})}\,|\,\mathfrak{Z}\,|^{\frac{1}{2}m-n-1}\,\{d\mathfrak{Z}\} = \frac{\rho_{m-2n-1}}{\rho_{m-2n+t-1}};$$

consequently, by (2), (121), (122), (123),

$$\begin{aligned}\lim_{\epsilon\to 0} \epsilon^{\frac{1}{2}n(m-n-1)} &\sum_{\mathfrak{S}[\mathfrak{G}]=\mathfrak{T}} \int_{H_0} e^{-\frac{1}{2}\pi\epsilon\sigma(\mathfrak{H}[\mathfrak{G}])}\, dv_H \\ &= \frac{2\rho_{m-2n-1}}{\rho_{r-n}\,\rho_{m-r-n}\,\rho_{m-n-1}}\, S^{\frac{1}{2}(m-n+1)}\, \rho(\mathfrak{S})\mu(\mathfrak{S}, \mathfrak{T}).\end{aligned} \tag{124}$$

Theorem 1 follows immediately from Lemma 30 and (124).

11. Proof of Theorem 2

Let $\mathfrak{G}^{(m,n)}$ be integral of rank n; then there exist a primitive $\mathfrak{F}^{(m,n)}$ and a non-singular integral $\mathfrak{C}^{(n)}$ such that $\mathfrak{G} = \mathfrak{F}\mathfrak{C}$, and $\mathfrak{C}$ is determined up to an arbitrary unimodular factor $\mathfrak{B}^{(n)}$ on the left side. If $\mathfrak{C}_1 = \mathfrak{B}\mathfrak{C}$, then $\mathfrak{C}^{-1}$ and $\mathfrak{C}_1^{-1}$ are left-equivalent. Once for all we choose a complete set Φ of integral $\mathfrak{C}$ with $|\mathfrak{C}| > 0$ and non-equivalent $\mathfrak{C}^{-1}$. The set Φ contains exactly one unimodular matrix; we may assume that this is the unit matrix. Let $\mu(\mathfrak{S}, \mathfrak{T})$ and $\nu(\mathfrak{S}, \mathfrak{T})$ be the quantities defined in (2) and (5). We shall suppose that $n < m$.

Lemma 33. *Let $\mathfrak{C}$ run over all matrices in Φ with integral $\mathfrak{T}[\mathfrak{C}^{-1}]$, then*

$$\mu(\mathfrak{S}, \mathfrak{T}) = \sum_{\mathfrak{C}} |\mathfrak{C}|^{n-m+1} \nu(\mathfrak{S}, \mathfrak{T}[\mathfrak{C}^{-1}]).$$

Proof. Let $\mathfrak{G}$ be a solution of $\mathfrak{S}[\mathfrak{G}] = \mathfrak{T}$, of rank n. Consider all real $\mathfrak{Y}^{(m,m-n)}$ such that the matrices $\mathfrak{Q}$, $\mathfrak{R}$ in $\mathfrak{S}[\mathfrak{G}, \mathfrak{Y}] = \mathfrak{W} = \begin{pmatrix} \mathfrak{T} & \mathfrak{Q} \\ \mathfrak{Q}' & \mathfrak{R} \end{pmatrix}$ lie in a given domain D; let D_0 be a fundamental domain in the $\mathfrak{Y}$-space, with respect to the transformations $\mathfrak{Y} \to \mathfrak{U}\mathfrak{Y}$, where $\mathfrak{U}$ is any element of $\Gamma(\mathfrak{S}, \mathfrak{G})$; then we have

$$\int_{D_0} \{d\mathfrak{Y}\} = S^{\frac{1}{2}} \rho(\mathfrak{S}, \mathfrak{G}) \int_D W^{-\frac{1}{2}} \{d\mathfrak{Q}\} \{d\mathfrak{R}\}. \tag{125}$$

Suppose that $\mathfrak{G} = \mathfrak{F}\mathfrak{C}$, with primitive $\mathfrak{F}$, and define $\mathfrak{S}[\mathfrak{F}, \mathfrak{Y}] = \mathfrak{W}_1 = \begin{pmatrix} \mathfrak{T}_1 & \mathfrak{Q}_1 \\ \mathfrak{Q}_1' & \mathfrak{R} \end{pmatrix}$; then $\mathfrak{S}[\mathfrak{F}] = \mathfrak{T}_1 = \mathfrak{T}[\mathfrak{C}^{-1}]$, $\mathfrak{Q}_1 = \mathfrak{Q}\mathfrak{C}^{-1}$, $\{d\mathfrak{Q}_1\} = C^{n-m}\{d\mathfrak{Q}\}$, $W_1 = C^{-2}W$. Since $\Gamma(\mathfrak{S}, \mathfrak{G}) = \Gamma(\mathfrak{S}, \mathfrak{F})$, we infer from (125) that $\rho(\mathfrak{S}, \mathfrak{G}) = C^{n-m+1}\rho(\mathfrak{S}, \mathfrak{F})$. The assertion follows readily from the definitions of $\mu(\mathfrak{S}, \mathfrak{T})$ and $\nu(\mathfrak{S}, \mathfrak{T})$.

Let q be a positive integer, and denote by Φ_q the subset of all $\mathfrak{C}$ in Φ with integral $q\mathfrak{C}^{-1}$.

Lemma 34. *Let $\mathfrak{G}^{(m,n)}$ be integral, then there exists in Φ_q a uniquely determined matrix $\mathfrak{C}$ such that $\mathfrak{G} \equiv \mathfrak{F}\mathfrak{C} \pmod q$, with primitive $\mathfrak{F}$.*

Proof. Let h be the rank of $\mathfrak{G}$. By the theory of elementary divisors, there exist two unimodular matrices $\mathfrak{U}_1^{(m)}$, $\mathfrak{U}_2^{(n)}$ and a rectangular diagonal matrix $\mathfrak{D}^{(m,n)} = [d_1, \cdots, d_n]$, with $1 \mid d_1 \mid \cdots \mid d_h$ and $d_k = 0$ $(h < k \leq n)$, such that $\mathfrak{G} = \mathfrak{U}_1\mathfrak{D}\mathfrak{U}_2$. Set $c_k = (d_k, q)$ $(k = 1, \cdots, n)$, $\mathfrak{C}_0^{(n)} = [c_1, \cdots, c_n]$; then $\mathfrak{C}_0$ has the elementary divisors $c_1, \cdots, c_n$ and $q\mathfrak{C}_0^{-1}$ is integral. We determine integers a_k $(k = 1, \cdots, m)$ such that $d_k \equiv a_k c_k \pmod q$ for $k \leq n$ and $\prod_{k=1}^{m} a_k \equiv 1 \pmod q$; this is possible, because of $n < m$. Choose a unimodular $\mathfrak{U}_0^{(m)} \equiv [a_1, \cdots, a_m] \pmod q$, and denote by $\mathfrak{F}_0$ the matrix of the first n columns in $\mathfrak{U}_1\mathfrak{U}_0$; then $\mathfrak{F}_0\mathfrak{C}_0 = \mathfrak{U}_1\mathfrak{U}_0\begin{pmatrix} \mathfrak{C}_0 \\ 0 \end{pmatrix} \equiv \mathfrak{U}_1\mathfrak{D} \pmod q$. Determine the unimodular matrix $\mathfrak{B}^{(n)}$ such that $\mathfrak{B}\mathfrak{C}_0\mathfrak{U}_2 = \mathfrak{C}$ lies in Φ, and define $\mathfrak{F}_0\mathfrak{B}^{-1} = \mathfrak{F}$; then $\mathfrak{F}$ is primitive, $q\mathfrak{C}^{-1}$ is integral and $\mathfrak{G} = \mathfrak{U}_1\mathfrak{D}\mathfrak{U}_2 \equiv \mathfrak{F}\mathfrak{C} \pmod q$.

If also $\mathfrak{G} \equiv \mathfrak{F}_1\mathfrak{C}_1 \pmod q$, with primitive $\mathfrak{F}_1$ and integral $q\mathfrak{C}_1^{-1}$, then $\mathfrak{F}_1\mathfrak{C}_1 \equiv \mathfrak{F}\mathfrak{C} \pmod q$; hence $\mathfrak{C}_1\mathfrak{C}^{-1}$ and $\mathfrak{C}\mathfrak{C}_1^{-1}$ are both integral; this proves that $\mathfrak{C}_1^{-1}$ and $\mathfrak{C}^{-1}$ are left-equivalent. Consequently, $\mathfrak{C}$ is uniquely determined in ϕ_q; q.e.d.

For any given $\mathfrak{C}$ in Φ_q, we denote by $A_q(\mathfrak{S}, \mathfrak{T}, \mathfrak{C})$ the number of modulo q incongruent integral solutions $\mathfrak{G}$ of $\mathfrak{S}[\mathfrak{G}] \equiv \mathfrak{T} \pmod{q}$, $\mathfrak{G} \equiv \mathfrak{F}\mathfrak{C} \pmod{q}$, with primitive $\mathfrak{F}$. By Lemma 34, we have

$$A_q(\mathfrak{S}, \mathfrak{T}) = \sum_{\mathfrak{C}} A_q(\mathfrak{S}, \mathfrak{T}, \mathfrak{C}), \tag{126}$$

where $\mathfrak{C}$ runs over ϕ_q and $A_q(\mathfrak{S}, \mathfrak{T})$ is the number of all solutions of $\mathfrak{S}[\mathfrak{G}] \equiv \mathfrak{T} \pmod{q}$. Let $\mathfrak{C}$ and a primitive $\mathfrak{F}^{(m,n)}$ be given and consider modulo q all primitive $\mathfrak{F}_1$ satisfying $\mathfrak{F}\mathfrak{C} \equiv \mathfrak{F}_1\mathfrak{C}_1 \pmod{q}$; their number $C_q(\mathfrak{C})$ does not depend on $\mathfrak{F}$. We remark that a given residue class $\mathfrak{F}_0^{(m,n)}$ modulo q contains a primitive $\mathfrak{F}$, if and only if $\mathfrak{F}_0$ is primitive modulo q; this means that all elementary divisors of $\mathfrak{F}_0$ are prime to q. On the other hand, let $B_q(\mathfrak{C})$ denote the number of modulo q incongruent primitive $\mathfrak{F}$ satisfying $\mathfrak{S}[\mathfrak{F}][\mathfrak{C}] \equiv \mathfrak{T} \pmod{q}$. Plainly,

$$A_q(\mathfrak{S}, \mathfrak{T}, \mathfrak{C}) = B_q(\mathfrak{C})/C_q(\mathfrak{C}), \tag{127}$$

$$B_q(\mathfrak{C}) = \sum_{\mathfrak{T}_1[\mathfrak{C}] \equiv \mathfrak{T} \pmod{q}} B_q(\mathfrak{S}, \mathfrak{T}_1), \tag{128}$$

where $B_q(\mathfrak{S}, \mathfrak{T}_1)$ is the number of primitive solutions $\mathfrak{F}$ of $\mathfrak{S}[\mathfrak{F}] \equiv \mathfrak{T}_1 \pmod{q}$ and $\mathfrak{T}_1$ runs modulo q over all integral symmetric matrices satisfying $\mathfrak{T}_1[\mathfrak{C}] \equiv \mathfrak{T} \pmod{q}$; let $D_q(\mathfrak{C})$ denote the number of these $\mathfrak{T}_1$.

Lemma 35. *Let p_0 be the product of all different prime factors of q, and let $c_1, \cdots, c_n$ be the elementary divisors of a given matrix $\mathfrak{C}$ in Φ_q; then*

$$C_q(\mathfrak{C}) \geqq \prod_{k=1}^{n} (c_k, p_0^{-1}q)^m.$$

If $c_n \mid p_0^{-1}q$, then $C_q(\mathfrak{C}) = |\mathfrak{C}|^m$.

Proof. It suffices to prove the lemma for $\mathfrak{C} = \mathfrak{C}_0 = [c_1, \cdots, c_n]$. If $\mathfrak{F}\mathfrak{C} \equiv \mathfrak{F}_1\mathfrak{C} \pmod{q}$, then $\mathfrak{F}_1 - \mathfrak{F} = \mathfrak{X}$ satisfies $\mathfrak{X}\mathfrak{C} \equiv 0 \pmod{q}$. The number of modulo q incongruent solutions $\mathfrak{X}^{(m,n)}$ of this congruence is $|\mathfrak{C}|^m$. Assume first that $q\mathfrak{C}^{-1} \equiv 0 \pmod{p_0}$; then each solution $\mathfrak{X}$ fulfills the congruence $\mathfrak{X} \equiv 0 \pmod{p_0}$. If $\mathfrak{F}$ is primitive, then $\mathfrak{F} + \mathfrak{X}$ is not necessarily primitive, but primitive modulo q, and we may choose $\mathfrak{X}$ in its residue class modulo q such that $\mathfrak{F} + \mathfrak{X} = \mathfrak{F}_1$ becomes primitive. This proves the second part of the assertion. If the condition $q\mathfrak{C}^{-1} \equiv 0 \pmod{p_0}$ is not satisfied, then we restrict the solutions $\mathfrak{X}$ of the congruence $\mathfrak{X}\mathfrak{C} \equiv 0 \pmod{q}$ by the further condition $\mathfrak{X} \equiv 0 \pmod{p_0}$; the remaining number of solutions is $\prod_{k=1}^{n} (c_k, p_0^{-1}q)^m$, whence the first part of the assertion.

Lemma 36. *Let $c_1, \cdots, c_n$ be the elementary divisors of the matrix $\mathfrak{C}$ in Φ_q; then*

$$D_q(\mathfrak{C}) \leqq \prod_{k=1}^{n} (c_k, q)^{n+1}.$$

If $c_n^2 \mid q$ and $D_q(\mathfrak{C}) \neq 0$, then $D_q(\mathfrak{C}) = |\mathfrak{C}|^{n+1}$.

Proof. By definition, $D_q(\mathfrak{C})$ is the number of $\mathfrak{T}_1$ modulo q satisfying $\mathfrak{T}_1[\mathfrak{C}] \equiv \mathfrak{T}$

(mod q). We may again assume that $\mathfrak{C} = \mathfrak{C}_0 = [c_1, \cdots, c_n]$. If $\mathfrak{T}_0$ is a given solution of the congruence, then the general solution is $\mathfrak{T}_1 \equiv \mathfrak{T}_0 + \mathfrak{Z}$ (mod q), with arbitrary symmetric $\mathfrak{Z}^{(n)}$ satisfying $\mathfrak{Z}[\mathfrak{C}] \equiv 0$ (mod q). Let $D_q(\mathfrak{C}) > 0$, then it follows that $D_q(\mathfrak{C}) = \prod_{k \leq l} (c_k c_l, q) \leq \prod_{k=1}^{n} (c_k, q)^{n+1} \leq \prod_{k=1}^{n} c_k^{\,n+1} = |\mathfrak{C}|^{n+1}$; the sign of equality being true, whenever $c_n^2 \mid q$; q.e.d.

LEMMA 37. *Let $q = p^a$ be a power of a given prime number p, and suppose that $4S^2$ is not divisible by q; then*

$$\beta_p(\mathfrak{S}, \mathfrak{T}) = q^{\frac{1}{2}n(n+1)-mn} B_q(\mathfrak{S}, \mathfrak{T}).$$

PROOF. Let p^b be the highest power of p dividing $2S$, then $a > 2b$. Consider a given primitive solution $\mathfrak{F}^{(m,n)}$ of $\mathfrak{S}[\mathfrak{F}] \equiv \mathfrak{T}$ (mod q), and let $\gamma(\mathfrak{F})$ be the number of modulo q incongruent primitive solutions which are congruent with $\mathfrak{F}$ modulo $p^{-b}q$. Denote by $d_1, \cdots, d_n$ the elementary divisors of $\mathfrak{S}\mathfrak{F}$ and determine unimodular $\mathfrak{U}_1^{(m)}$, $\mathfrak{U}_2^{(n)}$ such that $\mathfrak{U}_1\mathfrak{S}\mathfrak{F}\mathfrak{U}_2 = \binom{\mathfrak{D}}{0}$, $\mathfrak{D}^{(n)} = [d_1, \cdots, d_n]$; plainly, $d_1 \cdots d_n$ is a divisor of S, and $2d_k$ $(k = 1, \cdots, n)$ is not divisible by p^{b+1}. Set $\mathfrak{F}_1 = \mathfrak{F} + p^{-b}q\mathfrak{U}_1'\mathfrak{X}\mathfrak{U}_2^{-1}$, $\mathfrak{X} = \binom{\mathfrak{Y}}{\mathfrak{Z}}$, with integral $\mathfrak{Y}^{(n)}$ and $\mathfrak{Z}^{(m-n,n)}$; then $\mathfrak{F}_1 \equiv \mathfrak{F}$ (mod $p^{-b}q$) and

$$(\mathfrak{S}[\mathfrak{F}_1] - \mathfrak{S}[\mathfrak{F}])[\mathfrak{U}_2] \equiv p^{-b}q(\mathfrak{D}\mathfrak{Y} + \mathfrak{Y}'\mathfrak{D}) \qquad (\text{mod } pq).$$

Consequently, the congruence $\mathfrak{S}[\mathfrak{F}_1] \equiv \mathfrak{T}$ (mod pq) holds, if and only if $\mathfrak{Y} = (y_{kl})$ satisfies the conditions

$$(129) \qquad q^{-1}p^b(\mathfrak{T} - \mathfrak{S}[\mathfrak{F}])[\mathfrak{U}_2] \equiv \mathfrak{D}\mathfrak{Y} + \mathfrak{Y}'\mathfrak{D} \qquad (\text{mod } p^{b+1}).$$

Denote the left-hand member by $\mathfrak{W} = (w_{kl})$, then $w_{kl} \equiv 0$ (mod p^b), and (129) is the system of congruences $d_k y_{kl} + d_l y_{lk} \equiv w_{kl}$ (mod p^{b+1}), for $1 \leq k \leq l \leq n$. It follows that the number of solutions $\mathfrak{Y}$ modulo p^{b+1} is $p^{\frac{1}{2}n(n-1)}$ times the number of solutions $\mathfrak{Y}$ modulo p^b; since $\mathfrak{Z}$ is arbitrary, we infer that there are exactly $p^{\frac{1}{2}n(n-1)+n(m-n)}\gamma(\mathfrak{F})$ modulo pq incongruent primitive solutions of $\mathfrak{S}[\mathfrak{F}] \equiv \mathfrak{T}$ (mod pq) which are congruent with $\mathfrak{F}$ modulo $p^{-b}q$. Summing over all primitive residue classes $\mathfrak{F}$ modulo $p^{-b}q$, we infer that

$$p^{nm-\frac{1}{2}n(n+1)} B_q(\mathfrak{S}, \mathfrak{T}) = B_{pq}(\mathfrak{S}, \mathfrak{T}).$$

The assertion follows readily from the definition (6) of $\beta_p(\mathfrak{S}, \mathfrak{T})$.

Henceforth we suppose again that $2n + 2 < m$. Let q run over a sequence $q_1, q_2, \cdots$ of positive integers such that any positive integer divides q_k for all sufficiently large k. By Lemma 25, the limit

$$(130) \qquad \lim_{q\to\infty} q^{\frac{1}{2}n(n+1)-mn} A_q(\mathfrak{S}, \mathfrak{T}) = \lambda(\mathfrak{S}, \mathfrak{T}) = \prod_p \alpha_p(\mathfrak{S}, \mathfrak{T})$$

exists uniformly with respect to $\mathfrak{T}$ and its value is independent of the particular sequence Q. We denote by Q_0 any sequence Q with the additional property that q_k $(k = 1, 2, \cdots)$ contains all its prime factors at least to the m^{th} power.

Let a non-negative integer v be given, and define

$$\omega(v, q) = q^{\frac{1}{2}n(n+1)-mn} \sum_{c_n > v} A_q(\mathfrak{S}, \mathfrak{T}, \mathfrak{C}), \tag{131}$$

the sum extended over all $\mathfrak{C}$ in Φ_q with elementary divisor $c_n > v$.

LEMMA 38. *If ϵ is an arbitrarily small positive number, then there exists a positive integer $v_0 = v_0\,(\epsilon, m, S)$ depending only on ϵ, m, S such that $\omega(v, q) < \epsilon$, for any $v \geqq v_0$ and all sufficiently large q in Q_0.*

PROOF. By Lemmata 16 and 25,

$$B_q(\mathfrak{S}, \mathfrak{T}_1) \leqq A_q(\mathfrak{S}, \mathfrak{T}_1) < a_{32} q^{mn-\frac{1}{2}n(n+1)}, \tag{132}$$

for all sufficiently large q in Q_0, independent of $\mathfrak{T}_1$. Consequently, by (127), (128) and Lemmata 35, 36,

$$q^{\frac{1}{2}n(n+1)-mn} A_q(\mathfrak{S}, \mathfrak{T}, \mathfrak{C}) \leqq a_{32} D_q(\mathfrak{C})/C_q(\mathfrak{C}) \leqq a_{32} \prod_{k=1}^{n} (c_k, q)^{n+1}(c_k, p_0^{-1}q)^{-m}, \tag{133}$$

where p_0 is the product of the different prime factors of q and $c_1, \cdots, c_n$ are the elementary divisors of the matrix $\mathfrak{C}$ in ϕ_q. On the right-hand side of (131), we collect all terms with fixed elementary divisors $c_1, \cdots, c_n$ of $\mathfrak{C}$ and $c_n \mid q$; their number is equal to the index of the subgroup of all unimodular $\mathfrak{U}^{(n)}$ with integral $\mathfrak{C}_0\mathfrak{U}\mathfrak{C}_0^{-1}$ in the whole unimodular group, where $\mathfrak{C}_0 = [c_1, \cdots, c_n]$. Defining

$$\delta_k(q) = \sum_{1 \leqq d \mid q} (d, q)^{n+1}(d, p_0^{-1}q)^{-m}\, d^{2k-n-1} \qquad (k = 1, \cdots, n-1),$$

$$\delta(v, q) = \sum_{v < d \mid q} (d, q)^{n+1}(d, p_0^{-1}q)^{-m}\, d^{n-1} \prod_{p \mid d} (1 - p^{-1})^{1-n}, \tag{134}$$

we obtain, by Lemma 9, (131), (133),

$$\omega(v, q) \leqq a_{32}\, \delta(v, q) \prod_{k=1}^{n-1} \delta_k(q), \tag{135}$$

for all sufficiently large q in Q_0.

Plainly, $\delta_k(qq^*) = \delta_k(q)\delta_k(q^*)$, whenever $(q, q^*) = 1$. Moreover, for $q = p^a$ and $a \geqq m$, we have

$$\delta_k(q) = \sum_{l=0}^{a-1} p^{l(2k-m)} + p^m q^{2k-m} < (1 - p^{-2})^{-1} + p^{-m} < (1 - p^{-2})^{-2},$$

because of $m > 2n + 2 > 2k + 2$; hence

$$\prod_{k=1}^{n-1} \delta_k(q) < a_{33}. \tag{136}$$

In order to estimate $\delta(v, q)$, we multiply the general term of the sum in (134) by d^{-s}; we get a finite Dirichlet series $\delta(v, q, s)$. Plainly, $\delta(v, q, 0) = \delta(v, q)$; on the other hand, $\delta(v, q, s)$ is obtained from the Dirichlet series $\delta(0, q, s)$ by can-

celling the first v terms. We have $\delta(0, qq^*, s) = \delta(0, q, s)\delta(0, q^*, s)$, for $(q, q^*) = 1$. If $q = p^a$ and $a \geqq m$, then

$$\delta(0, q, s) = 1 + (1 - p^{-1})^{1-n} \left(\sum_{l=1}^{a-1} p^{l(2n-m-s)} + p^m q^{2n-m-s} \right)$$

$$< 1 + (1 - p^{-1})^{1-n}(p^{-2-s}(1 - p^{-2-s})^{-1} + p^{-a(2+s)})$$

$$< 1 + 2^n(p^{2+s} - 1)^{-1} < (1 - p^{-2-s})^{-2n}.$$

Hence $\delta(0, q, s) < \zeta^{2n}(s + 2)$, and it follows that the remainder term $\delta(v, q, s)$ of $\delta(0, q, s)$ tends to zero, for $v \to \infty$, uniformly in q and $s \geqq 0$. Consequently, $\delta(v, q) \to 0$, and the assertion follows from (135), (136).

LEMMA 39. *Let q run over a sequence Q and $\mathfrak{C}$ over all matrices in Φ with integral $\mathfrak{T}[\mathfrak{C}^{-1}]$, then*

$$\lim_{q\to\infty} q^{\frac{1}{2}n(n+1)-mn} \sum_{\mathfrak{C}} |\mathfrak{C}|^{n-m+1} B_q(\mathfrak{S}, \mathfrak{T}[\mathfrak{C}^{-1}]) = \lambda(\mathfrak{S}, \mathfrak{T}). \tag{137}$$

PROOF. Let v be a given positive integer. By (126), (127), (130), (131), we have

$$\lambda(\mathfrak{S}, \mathfrak{T}) = \lim_{q\to\infty} \left(\omega(v, q) + q^{\frac{1}{2}n(n+1)-mn} \sum_{c_n \leqq v} B_q(\mathfrak{C})/C_q(\mathfrak{C})\right), \tag{138}$$

where $\mathfrak{C}$ runs over all matrices in Φ_q with largest elementary divisor $c_n \leqq v$.

Let Q_0 denote the sequence of the m^{th} powers of the terms of the given sequence Q. Suppose that $(2Sv!)^3 \mid q$; this holds for all sufficiently large q in Q_0. Consider an integral symmetric matrix $\mathfrak{T}_1$ satisfying $\mathfrak{T}_1[\mathfrak{C}] \equiv \mathfrak{T} \pmod{q}$, with $c_n \leqq v$; then $\mathfrak{T}[\mathfrak{C}^{-1}] \equiv \mathfrak{T}_1 \pmod{c_n^{-2}q}$, hence $\mathfrak{T}[\mathfrak{C}^{-1}] = \mathfrak{T}^*$ is integral, and $\mathfrak{T}^* \equiv \mathfrak{T}_1 \pmod{c_n^{-2}q}$. Let p be a prime factor of q and suppose that p^c is the highest power of p dividing $c_n^{-2}q$. If $p^b \mid 2S$, then $c \geqq 3b$ and $c \geqq 1$, whence $c > 2b$. By Lemma 37, we infer that $B_q(\mathfrak{S}, \mathfrak{T}_1) = B_q(\mathfrak{S}, \mathfrak{T}^*)$. Consequently, by (128) and Lemmata 35, 36,

$$B_q(\mathfrak{C})/C_q(\mathfrak{C}) = B_q(\mathfrak{S}, \mathfrak{T}[\mathfrak{C}^{-1}])D_q(\mathfrak{C})/C_q(\mathfrak{C}) = |\mathfrak{C}|^{n-m+1}B_q(\mathfrak{S}, \mathfrak{T}[\mathfrak{C}^{-1}]). \tag{139}$$

Let $\epsilon > 0$ be given. By Lemma 38, we have $\omega(v, q) < \epsilon$, for any $v \geqq v_0(\epsilon, m, S)$ and all sufficiently large q in Q_0. In view of (138), (139), we obtain the inequality

$$\text{abs}\left(\lambda(\mathfrak{S}, \mathfrak{T}) - q^{\frac{1}{2}n(n+1)-mn} \sum_{c_n \leqq v} |\mathfrak{C}|^{n-m+1} B_q(\mathfrak{S}, \mathfrak{T}[\mathfrak{C}^{-1}])\right) < 2\epsilon, \tag{140}$$

for any given $v \geqq v_0$ and all sufficiently large q in Q_0; the summation is carried over all $\mathfrak{C}$ in Φ with $c_n \leqq v$ and integral $\mathfrak{T}[\mathfrak{C}^{-1}]$. On the other hand, it follows from (132) and Lemma 15 that

$$q^{\frac{1}{2}n(n+1)-mn} \sum_{c_n > v} |\mathfrak{C}|^{n-m+1} B_q(\mathfrak{S}, \mathfrak{T}[\mathfrak{C}^{-1}]) < \epsilon, \tag{141}$$

for all sufficiently large v and uniformly for q in Q_0. By (140), (141), the formula (137) of the assertion is proved for the sequence Q_0 instead of Q.

Let a non-negative integer v be given, and define

$$\omega(v, q) = q^{\frac{1}{2}n(n+1)-mn} \sum_{c_n > v} A_q(\mathfrak{S}, \mathfrak{T}, \mathfrak{C}), \tag{131}$$

the sum extended over all $\mathfrak{C}$ in Φ_q with elementary divisor $c_n > v$.

LEMMA 38. *If ϵ is an arbitrarily small positive number, then there exists a positive integer $v_0 = v_0(\epsilon, m, S)$ depending only on ϵ, m, S such that $\omega(v, q) < \epsilon$, for any $v \geqq v_0$ and all sufficiently large q in Q_0.*

PROOF. By Lemmata 16 and 25,

$$B_q(\mathfrak{S}, \mathfrak{T}_1) \leqq A_q(\mathfrak{S}, \mathfrak{T}_1) < a_{32} q^{mn-\frac{1}{2}n(n+1)}, \tag{132}$$

for all sufficiently large q in Q_0, independent of $\mathfrak{T}_1$. Consequently, by (127), (128) and Lemmata 35, 36,

$$q^{\frac{1}{2}n(n+1)-mn} A_q(\mathfrak{S}, \mathfrak{T}, \mathfrak{C}) \leqq a_{32} D_q(\mathfrak{C})/C_q(\mathfrak{C}) \leqq a_{32} \prod_{k=1}^{n} (c_k, q)^{n+1}(c_k, p_0^{-1}q)^{-m}, \tag{133}$$

where p_0 is the product of the different prime factors of q and $c_1, \cdots, c_n$ are the elementary divisors of the matrix $\mathfrak{C}$ in ϕ_q. On the right-hand side of (131), we collect all terms with fixed elementary divisors $c_1, \cdots, c_n$ of $\mathfrak{C}$ and $c_n \mid q$; their number is equal to the index of the subgroup of all unimodular $\mathfrak{U}^{(n)}$ with integral $\mathfrak{C}_0\mathfrak{U}\mathfrak{C}_0^{-1}$ in the whole unimodular group, where $\mathfrak{C}_0 = [c_1, \cdots, c_n]$. Defining

$$\delta_k(q) = \sum_{1 \leqq d \mid q} (d, q)^{n+1}(d, p_0^{-1}q)^{-m} d^{2k-n-1} \qquad (k = 1, \cdots, n-1),$$

$$\delta(v, q) = \sum_{v < d \mid q} (d, q)^{n+1}(d, p_0^{-1}q)^{-m} d^{n-1} \prod_{p \mid d} (1 - p^{-1})^{1-n}, \tag{134}$$

we obtain, by Lemma 9, (131), (133),

$$\omega(v, q) \leqq a_{32}\delta(v, q) \prod_{k=1}^{n-1} \delta_k(q), \tag{135}$$

for all sufficiently large q in Q_0.

Plainly, $\delta_k(qq^*) = \delta_k(q)\delta_k(q^*)$, whenever $(q, q^*) = 1$. Moreover, for $q = p^a$ and $a \geqq m$, we have

$$\delta_k(q) = \sum_{l=0}^{a-1} p^{l(2k-m)} + p^m q^{2k-m} < (1 - p^{-2})^{-1} + p^{-m} < (1 - p^{-2})^{-2},$$

because of $m > 2n + 2 > 2k + 2$; hence

$$\prod_{k=1}^{n-1} \delta_k(q) < a_{33}. \tag{136}$$

In order to estimate $\delta(v, q)$, we multiply the general term of the sum in (134) by d^{-s}; we get a finite Dirichlet series $\delta(v, q, s)$. Plainly, $\delta(v, q, 0) = \delta(v, q)$; on the other hand, $\delta(v, q, s)$ is obtained from the Dirichlet series $\delta(0, q, s)$ by can-

celling the first v terms. We have $\delta(0, qq^*, s) = \delta(0, q, s)\delta(0, q^*, s)$, for $(q, q^*) = 1$. If $q = p^a$ and $a \geqq m$, then

$$\begin{aligned}\delta(0, q, s) &= 1 + (1 - p^{-1})^{1-n} \left(\sum_{l=1}^{a-1} p^{l(2n-m-s)} + p^m q^{2n-m-s}\right)\\ &< 1 + (1 - p^{-1})^{1-n}(p^{-2-s}(1 - p^{-2-s})^{-1} + p^{-a(2+s)})\\ &< 1 + 2^n(p^{2+s} - 1)^{-1} < (1 - p^{-2-s})^{-2^n}.\end{aligned}$$

Hence $\delta(0, q, s) < \zeta^{2^n}(s + 2)$, and it follows that the remainder term $\delta(v, q, s)$ of $\delta(0, q, s)$ tends to zero, for $v \to \infty$, uniformly in q and $s \geqq 0$. Consequently, $\delta(v, q) \to 0$, and the assertion follows from (135), (136).

LEMMA 39. *Let q run over a sequence Q and $\mathfrak{C}$ over all matrices in Φ with integral $\mathfrak{T}[\mathfrak{C}^{-1}]$, then*

$$\lim_{q\to\infty} q^{\frac{1}{2}n(n+1)-mn} \sum_{\mathfrak{C}} |\,\mathfrak{C}\,|^{n-m+1} B_q(\mathfrak{S}, \mathfrak{T}[\mathfrak{C}^{-1}]) = \lambda(\mathfrak{S}, \mathfrak{T}). \tag{137}$$

PROOF. Let v be a given positive integer. By (126), (127), (130), (131), we have

$$\lambda(\mathfrak{S}, \mathfrak{T}) = \lim_{q\to\infty} \Big(\omega(v, q) + q^{\frac{1}{2}n(n+1)-mn} \sum_{c_n \leqq v} B_q(\mathfrak{C})/C_q(\mathfrak{C})\Big), \tag{138}$$

where $\mathfrak{C}$ runs over all matrices in Φ_q with largest elementary divisor $c_n \leqq v$.

Let Q_0 denote the sequence of the m^{th} powers of the terms of the given sequence Q. Suppose that $(2Sv!)^3 \mid q$; this holds for all sufficiently large q in Q_0. Consider an integral symmetric matrix $\mathfrak{T}_1$ satisfying $\mathfrak{T}_1[\mathfrak{C}] \equiv \mathfrak{T} \pmod{q}$, with $c_n \leqq v$; then $\mathfrak{T}[\mathfrak{C}^{-1}] \equiv \mathfrak{T}_1 \pmod{c_n^{-2}q}$, hence $\mathfrak{T}[\mathfrak{C}^{-1}] = \mathfrak{T}^*$ is integral, and $\mathfrak{T}^* \equiv \mathfrak{T}_1 \pmod{c_n^{-2}q}$. Let p be a prime factor of q and suppose that p^c is the highest power of p dividing $c_n^{-2}q$. If $p^b \mid 2S$, then $c \geqq 3b$ and $c \geqq 1$, whence $c > 2b$. By Lemma 37, we infer that $B_q(\mathfrak{S}, \mathfrak{T}_1) = B_q(\mathfrak{S}, \mathfrak{T}^*)$. Consequently, by (128) and Lemmata 35, 36,

$$B_q(\mathfrak{C})/C_q(\mathfrak{C}) = B_q(\mathfrak{S}, \mathfrak{T}[\mathfrak{C}^{-1}])D_q(\mathfrak{C})/C_q(\mathfrak{C}) = |\,\mathfrak{C}\,|^{n-m+1}B_q(\mathfrak{S}, \mathfrak{T}[\mathfrak{C}^{-1}]). \tag{139}$$

Let $\epsilon > 0$ be given. By Lemma 38, we have $\omega(v, q) < \epsilon$, for any $v \geqq v_0(\epsilon, m, S)$ and all sufficiently large q in Q_0. In view of (138), (139), we obtain the inequality

$$\operatorname{abs}\Big(\lambda(\mathfrak{S}, \mathfrak{T}) - q^{\frac{1}{2}n(n+1)-mn} \sum_{c_n \leqq v} |\,\mathfrak{C}\,|^{n-m+1} B_q(\mathfrak{S}, \mathfrak{T}[\mathfrak{C}^{-1}])\Big) < 2\epsilon, \tag{140}$$

for any given $v \geqq v_0$ and all sufficiently large q in Q_0; the summation is carried over all $\mathfrak{C}$ in Φ with $c_n \leqq v$ and integral $\mathfrak{T}[\mathfrak{C}^{-1}]$. On the other hand, it follows from (132) and Lemma 15 that

$$q^{\frac{1}{2}n(n+1)-mn} \sum_{c_n > v} |\,\mathfrak{C}\,|^{n-m+1} B_q(\mathfrak{S}, \mathfrak{T}[\mathfrak{C}^{-1}]) < \epsilon, \tag{141}$$

for all sufficiently large v and uniformly for q in Q_0. By (140), (141), the formula (137) of the assertion is proved for the sequence Q_0 instead of Q.

By Lemma 37, the quantity $q^{\frac{1}{2}n(n+1)-mn}B_q(\mathfrak{S}, \mathfrak{T}[\mathfrak{C}^{-1}])$ remains unchanged if we replace q by q^m, provided q is a multiple of $8S^3$. Since this condition is satisfied for all sufficiently large q in Q, the assertion of the lemma holds also for Q; q.e.d.

Let $\mathfrak{T}$ have the rank t; for the proof of Theorem 2 we may suppose that $\mathfrak{T} = \begin{pmatrix}\mathfrak{T}_0 & 0\\ 0 & 0\end{pmatrix}$, with non-singular $\mathfrak{T}_0^{(t)}$. Then $\mathfrak{T}[\mathfrak{C}^{-1}] = \mathfrak{T}^*$ has again the rank t, and we may define Φ such that also $\mathfrak{T}^* = \begin{pmatrix}\mathfrak{T}_0^* & 0\\ 0 & 0\end{pmatrix}$, with non-singular t-rowed $\mathfrak{T}_0^*$, and that $\mathfrak{T}_0^*$ is a fixed representative of its class. It follows that $\mathfrak{C} = \begin{pmatrix}\mathfrak{C}_1 & 0\\ * & *\end{pmatrix}$ and $\mathfrak{T}_0^*[\mathfrak{C}_1] = \mathfrak{T}_0$; hence $|\mathfrak{T}_0^*|$ is a factor of the given number $|\mathfrak{T}_0|$. On the other hand, there exist only a finite number of classes of non-singular integral t-rowed symmetric matrices with given determinant; consequently, $\mathfrak{T}_0^*$ and $\mathfrak{T}^*$ belong to a finite set.

By Theorem 1 and Lemmata 33, 39, we have

$$\lim_{q\to\infty} \sum_{\mathfrak{T}^*} (\nu(\mathfrak{S}, \mathfrak{T}^*) - q^{\frac{1}{2}n(n+1)-mn}B_q(\mathfrak{S}, \mathfrak{T}^*)) \sum_{\mathfrak{T}^*[\mathfrak{C}]=\mathfrak{T}} |\mathfrak{C}|^{n-m+1} = 0, \tag{142}$$

where q runs over a sequence Q; the inner summation extends over all $\mathfrak{C}$ in Φ satisfying $\mathfrak{T}^*[\mathfrak{C}] = \mathfrak{T}$, the outer summation over all integral $\mathfrak{T}^*$ for which this equation is solvable. Let h be the number of all prime factors of $|\mathfrak{T}_0|$, computed with their multiplicities. Suppose that Theorem 2 is true for all $\mathfrak{T}_0$ with less than h prime factors; this supposition is empty in case $h = 0$. In the outer sum on the left-hand side of (142), we have the term $\mathfrak{T}^* = \mathfrak{T}$ obtained for $\mathfrak{C}_1 = \mathfrak{E}$; for all other terms, finite in number, we have $\mathfrak{T}_0^*[\mathfrak{C}_1] = \mathfrak{T}_0$ with non-unimodular integral $\mathfrak{C}_1$, hence $|\mathfrak{T}_0^*|$ contains less than h prime factors. By Theorem 2 and Lemma 37, we infer that

$$\nu(\mathfrak{S}, \mathfrak{T}^*) = \lim_{q\to\infty} q^{\frac{1}{2}n(n+1)-mn}B_q(\mathfrak{S}, \mathfrak{T}^*) = \prod_p \beta_p(\mathfrak{S}, \mathfrak{T}^*), \tag{143}$$

for all $\mathfrak{T}^* \neq \mathfrak{T}$ in (142). Now it follows from (142) that (143) is also true for $\mathfrak{T}^* = \mathfrak{T}$; this completes the proof of Theorem 2.

12. Additional remarks

It is known that the formula (8) can be expressed as an identity in the theory of modular forms of degree n; cf. my papers: *Ueber die analytische Theorie der quadratischen Formen*, Ann. of Math. (2) 36, pp. 527–606 (1935); *Ueber die analytische Theorie der quadratischen Formen* II, Ann. of Math. (2) 37, pp. 230–263 (1936). Theorem 1 leads to a refinement of these results. Let $\mathfrak{G}^{(m,n)}$ be integral of rank h. We determine a primitive matrix $\mathfrak{F}^{(h,n)}$ and an integral matrix $\mathfrak{Q}^{(m,h)}$ such that $\mathfrak{G} = \mathfrak{Q}\mathfrak{F}$, the matrix $\mathfrak{Q}$ being uniquely determined up to an arbitrary unimodular factor on the right side. If $\mathfrak{S}[\mathfrak{Q}] > 0$, then we define

$$\tau(\mathfrak{S}, \mathfrak{G}) = \frac{\rho_m}{\rho_{m-h}} S^{-h/2} |\mathfrak{S}[\mathfrak{Q}]|^{\frac{1}{2}(m-h-1)} \frac{\rho(\mathfrak{S}, \mathfrak{Q})}{\rho(\mathfrak{S})};$$

we set $\tau(\mathfrak{S}, 0) = 1$, and $\tau(\mathfrak{S}, \mathfrak{G}) = 0$ otherwise. Moreover, let $\mathfrak{C}, \mathfrak{D}$ run over all h-classes ($h = 0, \cdots, n$) of coprime symmetric n-pairs defined in Lemma 5, and set

$$H(\mathfrak{S}, \mathfrak{C}, \mathfrak{D}) = e^{\frac{1}{4}\pi i m h} S^{-h/2} \mid \mathfrak{C}_0 \mid^{-m/2} \sum_{\mathfrak{L}(\mathrm{mod}\ \mathfrak{C}_0)} e^{-\pi i \sigma(\mathfrak{S}[\mathfrak{L}]\mathfrak{C}_0^{-1}\mathfrak{D}_0)},$$

whenever $\sigma(\mathfrak{S}[\mathfrak{L}]\mathfrak{C}_0\mathfrak{D}_0')$ is even for all integral $\mathfrak{L}^{(m,h)}$, and $H(\mathfrak{S}, \mathfrak{C}, \mathfrak{D}) = 0$ otherwise.

Let $\mathfrak{Z}^{(n)}$ be a complex symmetric matrix with positive imaginary part, and define

$$F(\mathfrak{S}, \mathfrak{Z}) = \sum_{\mathfrak{G}} \tau(\mathfrak{S}, \mathfrak{G}) e^{\pi i \sigma(\mathfrak{S}[\mathfrak{G}]\mathfrak{Z})},$$

the summation extended over a complete set of integral matrices $\mathfrak{G}^{(m,n)}$ which are non-associate, relative to the group $\Gamma(\mathfrak{S})$. Theorem 1 implies the identity

$$F(\mathfrak{S}, \mathfrak{Z}) = \sum_{\mathfrak{C},\mathfrak{D}} H(\mathfrak{S}, \mathfrak{C}, \mathfrak{D}) \mid \mathfrak{C}\mathfrak{Z} + \mathfrak{D} \mid^{-m/2} \tag{144}$$

$$(n \leqq r, n \leqq m - r, 2n + 2 < m);$$

the proof proceeds in the same way as in the first of my above-mentioned papers, and we omit it.

It may be seen from the reciprocity formula of the generalized Gaussian sums $H(\mathfrak{S}, \mathfrak{C}, \mathfrak{D})$ that $F(\mathfrak{S}, \mathfrak{Z})$ is a modular form, whenever $m - r$ is even. It is remarkable that in this case all coefficients $\tau(\mathfrak{S}, \mathfrak{G})$ of the Fourier series $F(\mathfrak{S}, \mathfrak{Z})$ are rational numbers.

For $n = 1$, the function $F(\mathfrak{S}, \mathfrak{Z})$ is a power series of the single variable $e^{\pi i z}$, namely

$$F(\mathfrak{S}, z) = 1 + \frac{\pi^{m/2}}{\Gamma\left(\dfrac{m}{2}\right)} S^{-\frac{1}{2}} \sum_{t=1}^{\infty} t^{\frac{1}{2}m-1} \mu(\mathfrak{S}, t) e^{\pi i t z},$$

with the definition (2) of $\mu(\mathfrak{S}, t)$. The corresponding Dirichlet series

$$\zeta(\mathfrak{S}, s) = \sum_{t=1}^{\infty} \mu(\mathfrak{S}, t) t^{-s}$$

has been investigated in my paper: *Ueber die Zetafunktionen indefiniter quadratischer Formen*. II, Math. Zeitschr. 44, pp. 398–426 (1938); in particular, I obtained a functional equation for $\zeta(\mathfrak{S}, s)$. If $m - r$ is even and $m > 4$, an independent proof of this result can be derived from (144).

On the other hand, the properties of $\zeta(\mathfrak{S}, s)$ are useful for the extension of Theorem 1 to the case $n = 1 \leqq r < m = 4$. By an application of Kloosterman's well known method, it is possible to extend Theorems 4 and 5 to this case, provided $\mathfrak{T} = t \neq 0$, and our former argument leading from there to Theorems 1 and 2 remains valid, with small modifications. Consequently, $\zeta(\mathfrak{S}, s)$ is a genus invariant. If $\mid \mathfrak{S} \mid$ is a square number, then also the value $t = 0$ presents no

difficulty. For irrational $|\mathfrak{S}|^{\frac{1}{2}}$ and $t = 0$, however, the estimates obtained by Kloosterman's method are too weak for a proof of Theorem 4; in this case we proceed in the following manner:

It is known that

$$\lim_{s \to 0} s\zeta(\mathfrak{S}^{-1}, s) = (-1)^{\frac{1}{2}(r+1)}\pi^{-1}S^{-\frac{1}{2}}\mu(\mathfrak{S}, 0) \qquad (r = 1, 3;\ m = 4), \tag{145}$$

$$\zeta(\mathfrak{S}^{-1}, 0) + \zeta(-\mathfrak{S}^{-1}, 0) = S^{-\frac{1}{2}}\mu(\mathfrak{S}, 0) - \mu(\mathfrak{S}^{-1}, 0) \qquad (r = 2;\ m = 4). \tag{146}$$

By (145), the quantity $\mu(\mathfrak{S}, 0)$ is a genus invariant, for $r = 1, 3$. In the remaining case $r = 2$, the expression $|\mathfrak{S}|^{-\frac{1}{2}}\mu(\mathfrak{S}, 0) - \mu(\mathfrak{S}^{-1}, 0)$ is a genus invariant, by (146). Consequently we have $\mu(\mathfrak{S}, 0) - \mu(\mathfrak{S}_0, 0) = |\mathfrak{S}|^{\frac{1}{2}}(\mu(\mathfrak{S}^{-1}, 0) - \mu(\mathfrak{S}_0^{-1}, 0))$, for any $\mathfrak{S}_0$ in the genus of $\mathfrak{S}$. On the other hand, it is easily seen that the quotient $\mu(\mathfrak{S}[\mathfrak{L}], 0)/\mu(\mathfrak{S}, 0)$ is rational, for any rational non-singular $\mathfrak{L}^{(4)}$. There exists a rational $\mathfrak{L}$ such that $\mathfrak{S}_0 = \mathfrak{S}[\mathfrak{L}]$; moreover, $\mathfrak{S}^{-1} = \mathfrak{S}[\mathfrak{S}^{-1}]$. Hence $\mu(\mathfrak{S}, 0) - \mu(\mathfrak{S}_0, 0)$ and $\mu(\mathfrak{S}^{-1}, 0) - \mu(\mathfrak{S}_0^{-1}, 0)$ are commensurable; since $|\mathfrak{S}|^{\frac{1}{2}}$ is irrational, it follows that $\mu(\mathfrak{S}, 0) = \mu(\mathfrak{S}_0, 0)$. This proves that $\mu(\mathfrak{S}, 0)$ is a genus invariant, for $r = 2$. Furthermore, my first proof of (8) can be extended to the case $t = 0$, $m = 4$, and then Theorem 1 follows in this case from (8) and the invariance of $\mu(\mathfrak{S}, 0)$.

A generalization of Theorem 1 holds good in any algebraic number field of finite degree, and an analogue exists for hermitian matrices.

Institute for Advanced Study